UC-NRLF
$B 530 840

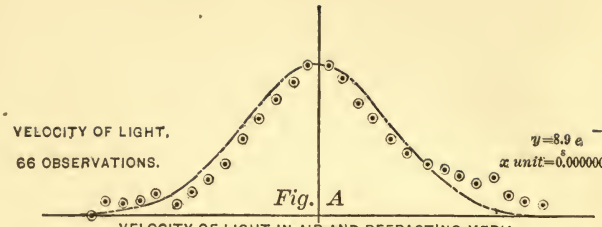
REESE LIBRARY

OF THE

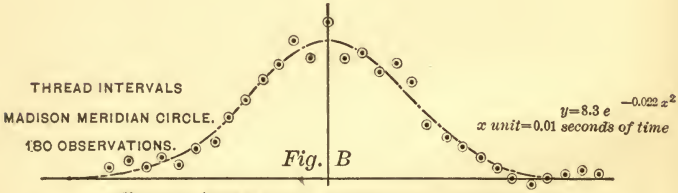
UNIVERSITY OF CALIFORNIA.

Class

THE METHOD OF LEAST SQUARES.

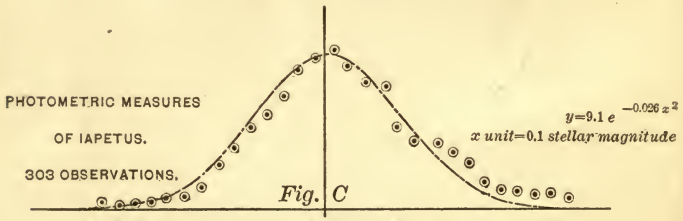


VELOCITY OF LIGHT IN AIR AND REFRACTING MEDIA.

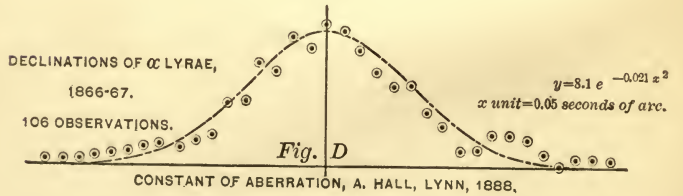


WASHBURN OBSERVATORY, UNPUBLISHED OBSERVATIONS.

ERRATUM. — *Fig. B.* The positive part of the axis of x should pass through the lowest plotted point. The relation of the plotted point to the curve is correctly represented.



ANNALS OF HARVARD COLLEGE OBSERVATORY, VOL. II., P. 252.



CONSTANT OF ABERRATION, A. HALL, LYNN, 1888.

TYPICAL ERROR CURVES.

AN ELEMENTARY TREATISE

UPON THE

METHOD OF LEAST SQUARES,

WITH

NUMERICAL EXAMPLES OF ITS APPLICATIONS.

BY

GEORGE C. COMSTOCK,

PROFESSOR OF ASTRONOMY IN THE UNIVERSITY OF WISCONSIN,
AND DIRECTOR OF THE WASHBURN OBSERVATORY.

BOSTON, U.S.A.:
PUBLISHED BY GINN & COMPANY.
1890.

Copyright, 1889, by
GEORGE C. COMSTOCK.
43202
All Rights Reserved.

Typography by J. S. Cushing & Co., Boston, U.S.A.
Presswork by Ginn & Co., Boston, U.S.A.

PREFACE.

THE following elementary treatment of the Method of Least Squares has grown out of my attempts to so present the subject to students of physics, astronomy, and engineering, that a working knowledge based upon an appreciation of its principles might be acquired with a moderate expenditure of time and labor.

Conceiving that the ultimate warrant for the legitimacy of the method itself is to be found in the agreement between the observed distribution of residuals and the distribution represented by the error curve, I have not scrupled to abandon altogether the analytical demonstrations of the equation of this curve and to present it as an empirical formula, representing the generalized experience of observers. The evidence in support of a formula of this kind is necessarily cumulative, and the few curves which are presented in illustration of the law of error are to be considered as samples of the kind of evidence which exists in great abundance. By abandoning the theoretical demonstrations, the student is freed from the embarrassments which are usually encountered at the threshold of the subject, and which in many cases cause it to appear as a mathematical puzzle whose analytical difficulties absorb the attention of the tyro to the complete exclusion of the purposes for which the analysis is conducted.

I have sought to give prominence to the distinction between accidental and systematic errors, and to insist upon the limi-

43202

tations which result from the difference between these two classes of error. To illustrate the principles of the text, I have made free use of numerical data and have arranged the computations in forms which experience has shown to be convenient for the purpose, with a view to their subsequent use by the student as models for his own computations.

In the preparation of these pages, I have consulted many, if not most, of the standard treatises upon the subject, but my indebtedness for suggestions and methods of treatment is principally to

FAYE, *Cours d'Astronomie de l'École Polytechnique.*

OPPOLZER, *Lehrbuch der Bahnbestimmung.*

WRIGHT, *Treatise on the Adjustment of Observations.*

G. C. C.

CONTENTS.

Section		Page
1.	Illustrative Problem	1
2.	Errors and Residuals	3
3.	The Distribution of Residuals	5
4.	The Error Curve	8
5.	The Principle of Least Squares	12
6.	Weights	16
7.	Normal Equations	19
8.	Non-Linear Observation Equations	21
9.	Formation and Solution of Normal Equations	23
10.	Numerical Example	29
11.	Conditioned Observations	38
12.	Probable Errors	45
13.	Probable Error of a Function of Observed Quantities	51
14.	Assignment of Weights; Rejection of Observations	54
15.	Empirical or Interpolation Formulæ	58
16.	Approximate Solutions	64
	Index to Formulæ	68

REESE LIBRARY OF THE UNIVERSITY OF CALIFORNIA.

THE METHOD OF LEAST SQUARES.

§ 1. Problem. To determine the coefficient of linear expansion of a certain bar of metal its length was determined at different temperatures by comparison with a standard of known length. The data furnished by the measures are *(Kohlrausch, Leitfaden der Physik)*:

Temperature.	Observed Length. mm.
20° C.	1000.22
40	1000.65
50	1000.90
60	1001.05

It is required to determine from these observations the amount of the expansion of the bar per degree Centigrade.

If c denote the required expansion, and l_0 the length of the bar when its temperature is 0° C., its length, l, at any other temperature, t, may be represented by the equation

$$l_0 + t \cdot c = l$$

By means of this equation the four observations recorded above are transformed into the following *observation equations*:

$$\left.\begin{aligned} (1)\quad & l_0 + 20\,c = 1000.22 \\ (2)\quad & l_0 + 40\,c = 1000.65 \\ (3)\quad & l_0 + 50\,c = 1000.90 \\ (4)\quad & l_0 + 60\,c = 1001.05 \end{aligned}\right\} \qquad (1)$$

Any two of these equations are sufficient to determine the values of l_0 and c, but the values derived from different pairs of equations will be different. Thus we may find from

Equations.	l_0	c
	mm.	mm.
(1) and (2)	999.79	+ 0.0215
(1) and (4)	999.80	.0208
(2) and (3)	999.65	.0250
(3) and (4)	1000.15	.0150
etc.	etc.	etc.

We are here presented with a problem of constant recurrence in the investigations and applications of physical science. In order to determine the values of certain quantities with a high degree of precision more measures or observations are made than are absolutely necessary, and these observations prove to be inconsistent among themselves, so that the resulting values of the unknown quantities depend upon the manner in which the data are combined. It is evident that all of the values above found for l_0 and c cannot be correct, and it is doubtful if any absolutely correct value can be derived from the data; but it is also apparent that the observations are not worthless and that any of the values above derived may be considered as approximations, more or less close, to the true values of the required quantities.

If we assume that the relation between the length of a bar and its temperature can be expressed by an equation of the form employed above, we must suppose that the discordances in the results are due to errors in the observations, and the problem then becomes:

To find from the observed data a set of results which shall be affected as little as possible by the errors of the data, or in more technical language, to find the most probable values of the unknown quantities.

We may establish in advance of any formal investigation of this problem certain principles to which its solution must conform. Thus,

(A) The adopted values of the quantities which are to be determined must be based upon *all* the data available. Only in exceptional cases, which will be considered hereafter, is it proper to omit or reject any observation or any known relation among the quantities.

(B) The adopted values must satisfy the observation equations as nearly as possible.

§ 2. **Errors and Residuals.** The expression *error of an observation* has been freely used in the preceding section, but it should be recognized that the amount of this error can rarely, if ever, be known, since this would imply an exact knowledge of the unknown quantities. We may, however, obtain approximate values of these errors from the adopted values of the quantities which were to be determined. Thus, if the values $l_0 = 999.79$, $c = +0.0215$ be substituted in equations (1), these become

(1) $1000.22 = 1000.22$ (3) $1000.87 = 1000.90$

(2) $1000.65 = 1000.65$ (4) $1001.08 = 1001.05$

The difference between the first and second members of any one of these equations is called the *residual* of that equation, and is approximately the error of the corresponding observation. The residuals which correspond to the several values of l_0 and c derived in § 1 are given below in tabular form.

$l_0 =$	999.79	999.80	999.65	1000.15
$c =$	+0.0215	+0.0208	+0.0250	+0.0150
$v =$	0.00	0.00	+0.07	−0.23
	.00	+ .02	.00	− .10
	− .03	+ .06	.00	.00
	+ .03	.00	− .10	.00

We may thus, for any assumed values of the unknown quantities, find a corresponding set of residuals, and the smaller these residuals are the closer is the probable approximation of the assumed, to the true values. Principle (B).

This statement, however, requires an important qualification to which we now proceed.

The errors with which any given series of observations is affected may be divided into two classes:

Accidental Errors, or those whose law of recurrence is such that in the long run they are as often positive as negative and

whose effect upon the mean of a great number of observations therefore differs but little from zero; and

Systematic Errors, or those which in the given series of observations do not thus tend to be eliminated from the mean. In the observations considered in § 1, an error of judgment by which the observer in a given case read the thermometer 0°.1 too high would probably be an accidental error, since it may be presumed that in the long run he would read it as often too low as too high, but if through a fixed habit of observing, the thermometer were always read too high this would be a systematic error, and the number of observations might be indefinitely increased without in the least diminishing its effect.

If the standard of length with which the bar was compared were an erroneous standard (*e.g.* 0.01 mm. too long), all of the observations would be affected with a systematic error due to this source, and the residuals would furnish no trace of this error, since they show only discordances among the observations, and not errors affecting all alike. The smallness of the residuals in any case, therefore, furnishes no guaranty that the observations and the results derived from them have not been vitiated by systematic errors.

The presence of errors of this class constitutes the greatest obstacle to the accurate determination of any set of quantities whose values are sought, and the ingenuity and skill of the observer or experimenter cannot be better employed than in avoiding or overcoming the effect of such errors. It therefore deserves especial notice that systematic errors can often be transformed into accidental errors by varying the methods of observation or the conditions under which the observations are made. Thus the possible systematic error of judgment in reading a thermometer, to which allusion was made above, may be transformed into an accidental error if several different persons take part in the observations, since it is hardly probable that they will all have a common, persistent error of judgment. The error due to using an erroneous standard of length may be changed into accidental error by employing a number of different standards, since it is not probable that these, con-

structed at different times and by different makers, will all have a common error of length. Considerations of this character serve to illustrate the great practical importance of varying the methods of determining any quantity whose value is desired with great precision. Multiplying observations by the same method and under similar circumstances serves only to diminish the effect of accidental errors and is useless beyond a certain limit, while varying the methods and the circumstances under which observations are made tends to eliminate errors of both kinds.

The principles here considered find their appropriate application in the selection of the methods by which any given set of unknown quantities is to be determined, but after the observations have been made, since they can, in general, furnish but little, if any, information in regard to their own systematic errors, these must be neglected and the reduction and discussion of the observations directed toward eliminating the effect of the accidental errors.

§ 3. **The Distribution of Residuals.** *Gauss*, a German mathematician, has shown by a course of analysis based upon the theory of probabilities that in any given series comprising a very large number of observations affected with accidental errors, the number of errors of a given magnitude, x, is a function of that magnitude. Thus, if x' and x'' denote any two errors, and y' and y'' the number of observations having the errors x' and x'' respectively, then

$$y' : y'' :: f(x') : f(x'')$$

The analytical expression for $f(x)$ obtained by Gauss is

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}, \qquad (2)$$

where e = base of the Naperian system of logarithms,
π = ratio of the circumference to the diameter of a circle,
h = a number whose value must be derived for each series of observations, but is constant for all the observations of that series.

The same expression for $f(x)$ has been derived by other mathematicians through different courses of analysis, but against all of these investigations objections of a theoretical character have been urged. Experience, however, shows that the actual distribution of residuals does follow this law, not with absolute accuracy, but to a remarkable degree of approximation. An excellent illustration of this distribution in the case of a comparatively small number of observations is afforded by a series of 66 determinations of the velocity of light made at Washington, in the year 1882.* By means of a revolving mirror the time required for the passage of a ray of sunlight from one terrestrial point to another was measured. The mean of the 66 determinations of this time interval was 24.827 millionths of a second. By subtracting this mean from each single determination a series of residuals will be obtained, and the number of residuals whose magnitude equals 1, 2, 3, etc. units may then be counted. In this way a fair approximation to the distribution of residuals represented by Gauss's law of error will be found; but as this law purports to represent the average distribution of a great number of errors, we shall obtain a better comparison between it and the actual distribution by the following device, to which we resort in order to increase the number of available residuals:

Let it be assumed that in any given set of observations the number of residuals of magnitude x is proportional to the number of residuals occurring between the limits $x - a$ and $x + a$, where a is a quantity which in strictness ought to be an infinitesimal, but which may be made a small finite quantity without appreciable error. In the present case we adopt as the unit in which the residuals are to be expressed, the thousand-millionth part of a second ($0^s.000000001$), and put a equal to two such units. Thus, from a series of 66 observations are derived the following numbers which represent the distribution of residuals which might be expected to occur in a much longer series.

---

* Velocity of Light in Air and Refracting Media. Bureau of Navigation, Navy Department, 1885, p. 187.

Residual.	No.	%	Residual.	No.	%
Less than −13.5	2	0.8	Greater than +13.5	0	0.0
Equal to −13.5	0	0.0	Equal to +13.5	2	0.8
−12.5	2	0.8	+12.5	2	0.8
−11.5	2	0.8	+11.5	3	1.2
−10.5	2	0.8	+10.5	6	2.3
− 9.5	3	1.2	+ 9.5	5	1.9
− 8.5	2	0.8	+ 8.5	6	2.3
− 7.5	4	1.6	+ 7.5	7	2.7
− 6.5	6	2.3	+ 6.5	8	3.1
− 5.5	8	3.1	+ 5.5	10	3.9
− 4.5	12	4.7	+ 4.5	12	4.7
− 3.5	15	5.8	+ 3.5	15	5.8
− 2.5	18	7.0	+ 2.5	17	6.6
− 1.5	21	8.2	+ 1.5	21	8.2
− 0.5	23	8.9	+ 0.5	23	8.9

The column headed % represents the number of residuals differing not more than half a unit from the magnitude given in the first column, expressed as a percentage of the whole number of residuals. Fig. A furnishes a graphical representation of this distribution, each percentage in the above table being represented by a point whose abscissa is the magnitude of the residual and whose ordinate is the percentage itself. The curve whose equation is

$$y = \frac{100\,h}{\sqrt{\pi}} e^{-h^2x^2} \qquad h = 0.158$$

is shown in the same figure, and a simple inspection of the curve shows that its ordinates represent very approximately the percentage of residuals of each magnitude. The coefficient h appears multiplied by the factor 100 in order that the ordinates may be represented as percentages.

Figs. B, C, D, represent the distribution of residuals in three other series of observations of different kinds, made at different places, by different observers, but all following the same law. The unit in which the residuals are expressed, unit of x, is stated with each figure, and the unit of y is in every case one per cent of the whole number of residuals.

The equations of the several curves shown in the figures are

almost identical, but the feature to which the student's attention is called is that the algebraic form of the equation is in each case

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

and not that h has approximately the same value in each curve. The numerical value of h depends upon the unit adopted for x, and these units having been chosen with reference to a convenient graphical representation of the residuals, the agreement in the several values of h must be regarded as purely artificial.

The series of observations represented in Fig. D is known to be affected with small systematic errors, and it will be noted that the distribution of the residuals is more irregular in this case than in any of the others. In each of the series represented in Figs. A and C, there are two residuals whose magnitudes are too great to be represented in the figures; and it is quite generally found that the actual number of very large residuals is slightly greater than the number given by the error curve. The illustrations here given are typical cases, and may serve to exemplify the statement made at the beginning of this section, that the actual distribution of residuals is found to follow Gauss's law of error, and in the following sections this law will be assumed as experimentally demonstrated, and from it will be derived the method of combining and discussing observations. The student will find it an instructive exercise to treat in a manner similar to that pursued above any series of observations to which he may have access, particularly his own observations, and thus lend additional weight to the experimental evidence which is here presented for his consideration.

§ **4. The Error Curve.** From the manner in which the ordinates of the points plotted in Figs. A, B, C, and D were derived, it will be apparent that these ordinates represent the number of residuals falling within certain chosen limits of error. Thus in Fig. A, 8.9 per cent of all the residuals lie between the

limits 0 and +1, 8.2 per cent between +1 and +2, etc., the interval within which the residuals are enumerated being in every case one unit. It is also evident that the number of residuals falling within any other interval, Δx, will depend upon the magnitude of this interval as well as upon the ordinate corresponding to it, and if Δx is taken sufficiently small the number of residuals will be proportional to the product $y \cdot \Delta x$. Geometrically considered, this product is the area included between the axis of x, the curve, and the two ordinates drawn through the extremities of Δx, and the number of residuals falling within the limits of Δx is therefore proportional to this area. We may, if we choose, make Δx an infinitesimal, and the area $y \cdot \Delta x$ and the corresponding number of residuals will then become indefinitely small, but by taking the sum of all the infinitesimal areas included between the limits $x = a$ and $x = b$, where a and b have any values whatever, we obtain the area of that part of the curve included between ordinates drawn at these limits. By a similar process of summation we obtain the number of residuals lying between a and b, and the number of residuals thus found must be proportional to the area, since this proportionality is true in every infinitesimal element included in the area.

In the following table, the function, A, represents the area of that part of the error curve included between ordinates whose abscissas are 0 and x, the argument of the table being the values of x for the particular error curve in whose equation $h = 1$; but the area included between 0 and x in the curve corresponding to any other value of h may be found from the same table, by using as the argument hx instead of x.

The area of that part of the curve lying between the limits a and b is represented by

$$A = \int_a^b y dx = \frac{h}{\sqrt{\pi}} \int_a^b e^{-h^2x^2} dx \qquad (3)$$

Let the variable in this expression be changed by putting $hx = t$, and the expression becomes

$$A = \frac{1}{\sqrt{\pi}} \int_{ha}^{hb} e^{-t^2} dt$$

FEDERAL LIBRARY

These expressions for A become identical if $h = 1$; hence, if the value of A be computed from the second integral for $h = 1$ and tabulated, we may find from this table the value of A corresponding to any other value of h by changing the limits a, b into ha and hb.

A remarkable property of the curve, which will be of use hereafter, may be readily obtained from the expression here found for A. If we make $a = 0$ and $b = \infty$, the limits of the integral become 0 and ∞ for all values of h, hence the area of that part of the curve included between $x = 0$ and $x = \infty$ is the same for all values of h, i.e., for every series of observations.

TABLE OF AREAS OF THE ERROR CURVE BETWEEN THE LIMITS 0 AND hx.

hx	A	Diff.	hx	A	Diff.	hx	A	Diff.
0.0	0.000		1.0	0.421		2.0	0.49766	
		56			19			85
0.1	.056		1.1	.440		2.1	.49851	
		55			15			56
0.2	.111		1.2	.455		2.2	.49907	
		53			12			36
0.3	.164		1.3	.467		2.3	.49943	
		50			9			23
0.4	.214		1.4	.476		2.4	.49966	
		46			7			14
0.5	.260		1.5	.483		2.5	.49980	
		42			5			9
0.6	.302		1.6	.488		2.6	.49989	
		37			4			4
0.7	.339		1.7	.492		2.7	.49993	
		32			3			3
0.8	.371		1.8	.495		2.8	.49996	
		27			1			2
0.9	.398		1.9	.496		2.9	.49998	
		23			2			1
1.0	.421		2.0	.498		3.0	.49999	
						∞	.50000	

If in any series of observations $n]_a^b$ denote the number of residuals whose magnitudes are included between the limits a and b, n the whole number of residuals in the series, and A_a,

A_b the values of A obtained from the table with the arguments ha, hb, then

$$n]_a^b = n(A_b \mp A_a)$$

since the ratio of $n]_a^b$ to n is equal to the ratio of the area of that part of the error curve which lies between the limits a and b, to the area of the whole curve, and this latter area is seen from the table to be always unity. The $-$ sign in this equation is to be used when a and b have like signs, and the $+$ when they have unlike signs. If the percentage of residuals between the limits a and b is required, it may be found by substituting 100 in place of n as the coefficient of $(A_b \mp A_a)$.

Thus from Fig. A we find for the series of observations there represented, $h^2 = 0.025$ and $h = 0.158$.

To find the distribution of residuals between the limits $-\infty \cdots -5$, $-5 \cdots -2$, $-2 \cdots +1$, $+1 \cdots +4$, $+4 \cdots +\infty$, we proceed as follows:

x	hx	A	$A_b \mp A_a$	$n]_a^b$ $= 66\,(A_b \mp A_a)$	Per cent.	Obs.
$-\infty$	$-\infty$	0.500				
			0.132	9	13.2	9
-5	-0.790	.368				
			.196	13	19.6	11
-2	-0.316	.172				
			.260	17	26.0	17
$+1$	$+0.158$	.088				
			.226	15	22.6	13
$+4$	$+0.632$	.314				
			.186	12	18.6	16
$+\infty$	$+\infty$	.500				
				66	100.0	66

The numbers in the column "Per cent" may be compared with the percentages given on page 7. The column "Obs." gives the actual number of residuals which occur in the given series between the limits here considered, and these numbers should be compared with the column "$n]_a^b$."

By the use of this table, the distribution of residuals in any series of observations for which the value of h is known may be compared with the theoretical distribution much more readily than by plotting a curve, and the student should in this way examine several series of observations. The method of determining h for any given series is contained in § 12.

§ 5. **The Principle of Least Squares.** The quantity h which appears in the equation of the error curve deserves especial attention. If in the equation

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

x be put equal to zero, the resulting value of y is $\frac{h}{\sqrt{\pi}}$. This is the maximum ordinate of the curve, and the value of this maximum ordinate varies directly as h. If those parts of the curve remote from the axis of y be considered, it will be found that the larger is h, the smaller are the values of y, since when x is a large quantity $e^{-h^2x^2}$ diminishes much more rapidly for increasing values of h than h itself increases. These relations between y and h correspond exactly to the criteria by which we estimate the precision of observations. If we compare two series of observations, I. and II., and find that in series I. the small errors are relatively more numerous (large values of y for small x's), and the large errors less numerous (small values of y for large x's), than in series II., we shall without hesitation call the observations of series I. more precise or accurate than those of series II.; and if required to assign definite meanings to the terms "more precise" and "less precise," we shall find difficulty in defining them in any other manner than by reference to the magnitude of the residuals. We therefore adopt as the *measure of precision* of any series of observations the value of h in the equation of its error curve; and having thus defined the term "precision," we are able to state two principles which are of general application in the discussion of observations.

Let the data furnished by each observation be expressed in the form of an observation equation (Equations 1, § 1), then: the best attainable values of the unknown quantities are those which,

(1) Distribute the residuals in accordance with the law of error, $y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$, and which,

(2) Make the value of h in the equation of the resulting error curve a maximum.

The first of these principles is indeed involved in the second, since if the residuals are not distributed in accordance with the law

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

there can be no value of h to be made a maximum. It is, however, advantageous to state (1) as a separate principle, since it affords a test of the presence of systematic errors in the data, which, though far from being a perfect criterion, is often convenient and is sometimes the only test available.

To justify the statement of (2), we resort to the following considerations: In accordance with A, § 1, we suppose that all of the data available is contained in the observation equations, and, B, § 1, we seek to satisfy all of these equations as nearly as possible. If the observations are free from systematic error, a supposition which must here be made, since we have no means of taking into account the effect of such errors, we may obtain by substituting in the observation equations any set of values which approximately satisfy them, a corresponding set of residuals which will be the errors of the observations, on the supposition that the substituted values were the true values of the unknown quantities. If these residuals are plotted in an error curve, they will furnish a numerical measure of the precision h, assigned to the observations by this set of values, and out of all possible sets of values of the unknown quantities that set which assigns the maximum precision to the observations will be entitled to the greatest degree of confidence; for if it were otherwise, we should have no reason for preferring a set of values which exactly satisfied all of the equations to a set which did not satisfy them.

It is, of course, true that subsequent observations may furnish a better determination of the unknowns, and that the values thus found will not assign to the earlier observations as high a degree of precision as did the erroneous values obtained

from these observations alone, but this subsequent determination is based upon additional evidence, and the problem with which we are concerned is not to obtain the best possible values of the unknown quantities, but the best values which can be derived from the data in our possession.

Assuming, then, the validity of (2), we proceed to transform it into an expression more convenient for practical use, and for this purpose we resort to the following property of the error curve, which may be approximately verified by actual measurement from any plotted curves, Figs. A, B, C, D.

If the error curve be divided into a great number of parts by drawing equidistant ordinates throughout its whole extent, and the areas of the several parts into which the curve is thus divided be each multiplied by the square of the abscissa of its middle point, the sum of all these products will equal $\frac{1}{2h^2}$.

The analytical expression for the process above described is

$$\int_{-\infty}^{+\infty} x^2 y\, dx \quad \text{or} \quad \frac{2h}{\sqrt{\pi}} \int_0^{\infty} x^2 e^{-h^2x^2}\, dx$$

Put $hx = t$, and this integral becomes

$$\frac{2}{h^2\sqrt{\pi}} \int_0^{\infty} t^2 e^{-t^2}\, dt = \frac{1}{2h^2}$$

For the method of obtaining the value of the last integral, see *Newcomb's Calculus*, Articles 169, 176.

The area of each of the parts into which the curve was divided is proportional to the number of residuals occurring between the limiting ordinates of the part; thus, let A denote the area of the part, N the corresponding number of residuals, and n and a the whole number of residuals and the whole area of the curve respectively; then

$$A : N :: a : n$$

but from the table in § 4, $a = 1$, whence

$$A = \frac{N}{n} \text{ and } Ax^2 = \frac{Nx^2}{n}$$

Since N denotes the number of x's falling within the given infinitesimal part, A, of the curve, Nx^2 is equal to the sum of the squares of the x's (residuals) whose magnitudes fall between the limiting ordinates of A, and taking the sum of all the Ax^2's we obtain

$$\Sigma Ax^2 = \frac{1}{n}\Sigma x^2$$

i.e., is equal to the mean of the squares of all the residuals. It is customary to represent the sum of the squares of the residuals by the symbol $[vv]$, v standing for any residual, and the [ ] denoting the sum of all quantities of the kind written within them.

Comparing this result with the one obtained above, we have

$$\frac{[vv]}{n} = \frac{1}{2h^2} \tag{4}$$

from which it appears that the relation between h and the sum of the squares of the residuals is such that when h is a maximum, $[vv]$ is a minimum, and principle (2) may be restated as follows:

The most probable values of the unknown quantities are those which make the sum of the squares of the residuals a minimum.

From this principle has been derived the name *Method of Least Squares*, which is commonly applied to that body of principles which treats of the combination and discussion of observed data.

We have arrived at this principle from a consideration of that class of cases in which the quantity observed is a function of two or more unknown quantities whose values are to be obtained from the observations. This obviously includes the case of a single quantity, x, whose value is directly measured; and it will be advantageous to apply the principle of least squares to this case. The observation equations are here of the simplest possible form.

$$x = m_1$$
$$x = m_2$$
$$x = m_3$$
$$\text{etc.}$$

where m denotes an observed value of x.

If x_0 denote any assumed value of x, the residuals obtained by substituting it in these equations will be

$$\begin{aligned} v_1 &= m_1 - x_0 \\ v_2 &= m_2 - x_0 \\ v_3 &= m_3 - x_0 \\ \cdots &\quad \cdots \quad \cdots \\ v_n &= m_n - x_0 \end{aligned}$$

and $[vv] = (m_1 - x_0)^2 + (m_2 - x_0)^2 + (m_3 - x_0)^2 \cdots + (m_n - x_0)^2$.

The value of x_0 which will make $[vv]$ a minimum is found from

$$\frac{d[vv]}{dx_0} = 0 = -2(m_1 - x_0) - 2(m_2 - x_0) - 2(m_3 - x_0) \cdots - 2(m_n - x_0)$$

but this equation is equivalent to

$$x_0 = \frac{m_1 + m_2 + m_3 + \cdots + m_n}{n} \equiv \frac{[m]}{n}$$

and it thus appears that the universal practice of taking the arithmetical mean of all the measures of a single quantity as the best value of that quantity, is a particular case under the more general method of least squares.

§ **6. Weights.** It frequently happens that the circumstances under which an observation was made lead the observer to distrust its accuracy, while other causes give him increased confidence in another observation. Observations which thus differ in quality are said to have different weights, the weight being a numerical measure of the quality, and these weights should be taken into account in combining the observations.

Let us suppose two series of observations made upon the same unknown quantity, in one of which the observations are of different quality and entitled to different degrees of confidence, while in the other the observations are all equally good, but each of them entitled to less confidence than the poorest observation of the first series. By taking the mean of a number of observations of this second series, a more reliable value of the unknown quantity may be obtained than any single

observation of the series can furnish, and by properly choosing the number of observations to be included in the mean, a value entitled to as much confidence as any observation of the first series may be found. This number of observations of the second series whose mean is entitled to as much confidence as a single observation of the first series, is called the *weight* of the equivalent observation in the first series; and, obviously, the better an observation, the greater is its weight. These weights furnish no information about the absolute precision of the observations, but express only their relative excellence as compared with each other; hence, if p_1, p_2, p_3, etc., be the weights of any observations, kp_1, kp_2, kp_3, etc., where k is any constant, will express these weights equally well, since it is the ratios of the weights, and not their absolute values, which are of importance.

To exhibit the manner in which these weights are to be employed, let us recur to the data of § 1, and suppose that those observations were made under such conditions that the first one has a weight 1, the second 2, the third 3, and the fourth 4. In accordance with the definition of weights, this is equivalent to supposing a second series of observations of uniform excellence, such that the first of the actual observations can be replaced by one observation of this series, which must of course be numerically the same as the observation which it replaces; the second real observation may be replaced by two numerically equal observations of the second series; the third by three, etc. Each of these substituted observations will furnish an equation precisely like those given in § 1, and when the sum of the squares of the residuals is formed, we shall obtain

$$v_1^2+(v_2^2+v_2^2)+(v_3^2+v_3^2+v_3^2)+(v_4^2+v_4^2+v_4^2+v_4^2)\equiv[pvv]$$

The symbol $[pvv]$, which is adopted as an abbreviation for this expression, is equivalent, numerically, to the sum obtained by multiplying the square of each actual residual by the weight of the corresponding observation and adding the products, and it is evident that this $[pvv]$ bears the same relation to the

substituted observations that $[vv]$ bore to the actual observations in the case of equal weights, which was considered in the preceding section. The principle there obtained may therefore be generalized as follows:

The most probable values of the unknown quantities are those which make the sum of the weighted squares of the residuals, $[pvv]$, *a minimum.*

Let the student show, as in the preceding section, that when this principle is applied to the case of observations of unequal weight made upon a single unknown quantity, it gives as the most probable value of that quantity

$$x_0 = \frac{[pm]}{[p]}$$

As an example of the application of weights, we select the following observations of the time of ending of the transit of Mercury of May 6, 1878, which were observed by different observers in the city of Washington. These observers were provided with telescopes of different sizes and magnifying powers, and differed among themselves in point of experience and skill, so that their observed times of last contact are not entitled to equal confidence. The weights assigned to the several observations represent the judgment of the computer with respect to their relative excellence. (Washington Observations, 1876, Appendix II., page 55.)

Observed Time.	p	pm	
$5^h\ 38^m\ 23^s$	1	23^s	$[pm] = 318$
37. 55	0	0	
38 10	1	10	$[p] = 16$
38 26	3	78	
38 21	2	42	$\frac{[pm]}{[p]} = 19.9$
38 18	2	36	
38 19	3	57	
38 21	2	42	
38 15	2	30	

Weighted mean of the observations $= 5^h\ 38^m\ 19^s.9$.

§ 7. **Normal Equations.** We have now to show how the principle of least squares is to be applied in determining the values of a set of unknown quantities, and in order to fix the ideas as definitely as possible, let it be supposed that there are three of these quantities, x, y, z, which are connected with each one of a set of observed quantities, n, by the relation

$$ax + by + cz = n$$

where a, b, and c are numerical coefficients whose values are supposed known in each equation.

From a series of more than three observed values of n, the most probable values of x, y, z are to be obtained by means of the relation $[pvv]$ = a minimum.

It is not to be presumed that these values when found will exactly satisfy all the equations, and make $[pvv] = 0$, but we shall find from each equation a residual v, so that strictly the observation equations should be written

$$\begin{aligned} a_1x + b_1y + c_1z - n_1 &= v_1 \qquad p_1 \\ a_2x + b_2y + c_2z - n_2 &= v_2 \qquad p_2 \\ a_3x + b_3y + c_3z - n_3 &= v_3 \qquad p_3 \\ \text{etc.} \qquad \text{etc.} &\qquad \text{etc.} \end{aligned}$$

The symbols p_1, p_2, p_3 represent the weights assigned to the observed values, n_1, n_2, n_3, etc.

By the ordinary rule for determining a minimum of a function of several variables, the condition $[pvv]$ = a minimum, furnishes the three equations

$$\frac{d[pvv]}{dx} = 0 \qquad \frac{d[pvv]}{dy} = 0 \qquad \frac{d[pvv]}{dz} = 0$$

and in order to obtain these derivatives we form from the observation equations

$$\left.\begin{aligned} [pvv] = &\{a_1\sqrt{p_1}x + b_1\sqrt{p_1}y + c_1\sqrt{p_1}z - \sqrt{p_1}n_1\}^2 \\ &+ \{a_2\sqrt{p_2}x + b_2\sqrt{p_2}y + c_2\sqrt{p_2}z - \sqrt{p_2}n_2\}^2 \\ &+ \{a_3\sqrt{p_3}x + b_3\sqrt{p_3}y + c_3\sqrt{p_3}z - \sqrt{p_3}n_3\}^2 \\ &\text{etc.} \qquad \text{etc.} \qquad \text{etc.} \end{aligned}\right\} \qquad (5)$$

The derivative of this expression with respect to x is

$$\left.\begin{aligned}\frac{d[pvv]}{dx} &= 2a_1\sqrt{p_1}\{a_1\sqrt{p_1}x + b_1\sqrt{p_1}y + c_1\sqrt{p_1}z - \sqrt{p_1}n_1\}\\ &+ 2a_2\sqrt{p_2}\{a_2\sqrt{p_2}x + b_2\sqrt{p_2}y + c_2\sqrt{p_2}z - \sqrt{p_2}n_2\}\\ &+ 2a_3\sqrt{p_3}\{a_3\sqrt{p_3}x + b_3\sqrt{p_3}y + c_3\sqrt{p_3}z - \sqrt{p_3}n_3\}\\ &\quad \text{etc.} \qquad \text{etc.} \qquad \text{etc.} \qquad = 0\end{aligned}\right\} \quad (6)$$

Let this expression be expanded, divided by 2 and simplified by the introduction of [ ] to denote the sum of all terms like those placed within them (all terms standing in the same vertical column), and it becomes the first of the following group of

NORMAL EQUATIONS.

$$\left.\begin{aligned}[paa]x + [pab]y + [pac]z - [pan] &= 0\\ [pab]x + [pbb]y + [pbc]z - [pbn] &= 0\\ [pac]x + [pbc]y + [pcc]z - [pcn] &= 0\end{aligned}\right\} \quad (7)$$

The second and third of these equations are derived in precisely the same manner as the first from the conditions

$$\frac{d[pvv]}{dy} = 0 \qquad \frac{d[pvv]}{dx} = 0$$

These equations are equal in number to the unknown quantities, and their solution will in general furnish a determinate set of values for these quantities which will be the most probable values, since the normal equations include all of the data furnished by the observations and have been so derived as to satisfy the principle of least squares.

Equations (6) furnish a rule which is frequently given for the formation of normal equations. To obtain the first normal equation, multiply each observation equation by the product of its weight into the coefficient of x which occurs in it, and take the sum of all the resulting equations. The other normals are similarly obtained from the weights and the coefficients of y, z, etc., having due regard to the algebraic signs of the quantities in the several multiplications and divisions. This method

is occasionally convenient, but in general the method of forming normal equations given in § 9 will be found less laborious.

The symmetrical manner in which the coefficients of the normal equations are disposed should be especially noted, since this considerably diminishes the labor of their formation. The first coefficient in the second equation is the same as the second coefficient in the first equation, and generally the m^{th} coefficient in the n^{th} equation is the same as the n^{th} coefficient in the m^{th} equation.

Let the student form normal equations from the observation equations contained in § 1, assuming that those equations have equal weights.

§ 8. Non-Linear Observation Equations. In all of the preceding investigation, it has been tacitly assumed that the relation of the observed to the unknown quantities can be expressed by an equation of the first degree; but cases in which this relation is of a much more complicated character are not uncommon, and a method of applying the principle of least squares to these cases is required. For the sake of simplicity, this method will be derived for the case of two unknown quantities, but the process is perfectly general and can readily be extended to any other number of unknowns.

Let x and y be any two quantities which have not been directly measured but which are connected with an observed quantity, m, by the relation

$$f(x, y, m) = 0$$

which represents any equation whatever existing between x, y, and m. Let x_0 and y_0 denote approximate values of x and y, such that

$$x = x_0 + \Delta x \qquad y = y_0 + \Delta y$$

Δx and Δy being the corrections which must be added to x_0 and y_0 in order to obtain the most probable values of x and y. We may, for the present, suppose that x_0 and y_0 are mere guesses at the values of x and y, and we may test their correctness by substituting their numerical values in the equation

$$f(x, y, m) = 0$$

which corresponds to each observed value of m. If every such equation were exactly satisfied by these values we should infer that x_0 and y_0 were the most probable values of x and y. It cannot be expected that this perfect agreement will ever be found in practice, but from each observation equation a residual, v, will be found, due partly to the errors of the observations and partly to Δx and Δy.

If in the equation $f(x, y, m) = 0$ we substitute for x and y $x_0 + \Delta x$, $y_0 + \Delta y$, and develop the expression by *Taylor's Formula*, remembering that $f(x_0, y_0, m)$ is the residual found by substituting numerical values of x_0, y_0 in the several observation equations, we have

$$f(x_0 + \Delta x, y_0 + \Delta y, m) = 0 = v + \frac{df}{dx_0}\Delta x + \frac{df}{dy_0}\Delta y + \cdots$$

If numerical values of v, $\frac{df}{dx_0}$, $\frac{df}{dy_0}$ be introduced into this equation, it becomes

$$a \cdot \Delta x + b \cdot \Delta y + \overset{v}{n} = 0$$

Each observation equation may thus be made to furnish a linear equation involving Δx and Δy, and these equations may be treated by the method of § 7. It must, however, be remembered that in the above development by *Taylor's Formula* we have retained only the first three terms of an infinite series, and if the approximate values x_0, y_0 are not so nearly the most probable values that the squares and higher powers of Δx and Δy are inappreciable, the development and the solution based upon it are inaccurate. On this account, it is seldom advantageous to make a least square solution for the unknown quantities until very approximate values of them have been found. These values will usually be obtained from the solution of a small number of the observation equations.

The transformation of the observation equations by the introduction of corrections to assumed values of the unknowns is often advantageous even when the original equations are of the first degree, especially if the original quantities were of very different magnitudes. Thus, in the problem of § 1, the observation equations are of the form

$$f(l_0, c, m) = l_0 + tc - m = 0$$

in which c is a very small quantity while l_0 is approximately 1000. If we put

$$l_0 = 1000 + \Delta l_0 \qquad c = 0 + \Delta c$$

we have $\frac{df}{dl_0} = 1 \quad \frac{df}{dc} = t \quad v = f(1000, 0, m) = 1000 - m$

and the equations are transformed into

$$\Delta l_0 + 20\,\Delta c - 0.22 = 0$$
$$\Delta l_0 + 40\,\Delta c - 0.65 = 0$$
$$\Delta l_0 + 50\,\Delta c - 0.90 = 0$$
$$\Delta l_0 + 60\,\Delta c - 1.05 = 0$$

By this transformation the numerical operations involved in forming and solving the normal equations are much simplified through the substitution of small numbers in the place of large ones.

§ 9. **Formation and Solution of the Normal Equations.** If the number of unknown quantities is greater than two, and especially if the number of observations is large, the numerical computation of a set of normal equations is a laborious process, and one in which errors are almost certain to occur unless special precautions are taken to guard against them. The method of forming these equations presented in this section has been developed with special reference to facilitating the numerical operations and obtaining the normals with the least expenditure of labor consistent with the requisite accuracy, and although some of the processes may seem at first sight unnecessary and cumbrous, a little experience in their use, or in their neglect, will convince the student that they are in the long run labor-saving devices.

Let each observation equation be written out and arranged in tabular form, as in the following example. In order that these equations should furnish a good determination of the unknowns, x, y, z, it is necessary that the coefficients of these quantities should present a considerable range in their values

in the several equations. Thus, if all the coefficients of x were alike, all the coefficients of y equal each to each, etc., the equations would be absolutely indeterminate since we should have several unknown quantities involved in a single equation many times repeated, and if the coefficients approximate to this equality the equations will be approximately indeterminate, and will furnish unreliable values of the unknowns. If, therefore, several observations have been made under similar conditions, and furnish equations which are nearly identical, these will be nearly equivalent to a repetition of the same equation and it will be permissible to take their sum, having regard to their respective weights, and treat it as a single observation equation with a weight equal to the sum of the weights of the observations.

Having thus reduced the number of equations as far as possible, each equation should be multiplied by the square root of its weight as was done, § 7, in obtaining the form of the normal equations. By this multiplication the weights will be completely taken into account and will require no further attention. Let the *weighted equations* thus obtained be represented by

$$a_1x + b_1y + c_1z + n_1 = 0$$
$$a_2x + b_2y + c_2z + n_2 = 0$$
$$a_3x + b_3y + c_3z + n_3 = 0$$
$$\text{etc.} \quad \text{etc.} \quad \text{etc.}$$

It will usually facilitate the formation of the normals to so transform these equations that no number greater than 1 shall occur in any of them. This can always be done by introducing new unknown quantities and dividing each equation by some constant number, usually some power of 10. Thus in the case of the two equations

$$5x + 71y - 63 = 0$$
$$0.9x - 193y + 93 = 0$$

let each equation be divided by 100 and put

$$\tfrac{5}{100}x = u \qquad \tfrac{193}{100}y = w;$$

the equations are thus transformed into

$$1.000\,u + 0.368\,w - 0.630 = 0$$
$$0.180\,u - 1.000\,w + 0.930 = 0$$

The solution of these equations will furnish values of u and w from which x and y may be found by the relations

$$x = 20\,u \qquad y = \tfrac{100}{193}\,w$$

The purpose of this transformation is to simplify the subsequent numerical work by reducing the numbers involved to an approximate equality.

Every coefficient which appears in the normal equations is the sum of a series of products of two quantities, thus

$$[aa] = a_1a_1 + a_2a_2 + a_3a_3 + \cdots$$
$$[ab] = a_1b_1 + a_2b_2 + a_3b_3 + \cdots$$
$$[bn] = b_1n_1 + b_2n_2 + b_3n_3 + \cdots$$

These products may be formed by the aid of *Crelle's* multiplication tables * supplemented by a table of squares of numbers for the $[aa]$, $[bb]$, etc. In case *Crelle's* tables are not available, the products may be formed by logarithms or much more rapidly by the following method due to *Bessel.* Form for each equation the sums $a + b$, $a + c$, $b + c$, etc., for every pair of numbers contained in the equation; then since

$$ab = \tfrac{1}{2}\{(a+b)^2 - aa - bb\}$$

we have

$$\left.\begin{aligned} [ab] &= \tfrac{1}{2}\{[(a+b)^2] - [aa] - [bb]\} \\ [bc] &= \tfrac{1}{2}\{[(b+c)^2] - [bb] - [cc]\} \\ &\text{etc.} \qquad\quad \text{etc.} \qquad\quad \text{etc.} \end{aligned}\right\} \qquad (8)$$

The $[aa]$, $[bb]$, $[cc]$ are coefficients in the normal equations and must be computed in any case, and the formation of $[ab]$, $[bc]$, etc., therefore requires for each coefficient only the single additional quantity $[(a+b)^2]$, $[(b+c)^2]$, and presents the very great advantage that these quantities can be obtained

* Crelle, Rechentafeln, Berlin. These tables give the products of all numbers up to 1000×1000, and are of very general utility.

from a table of squares, and being all positive numbers no attention need be paid to the signs after the sums $a+b$, $b+c$, etc., have been formed.

No method of computation can furnish a guaranty against the commission of numerical errors, and it is therefore desirable to test the computation from time to time to ascertain if such errors have occurred. To secure such a test or "check," as it is called, we introduce the following auxiliary quantities, one for each observation equation:

$$\left.\begin{array}{l} s_1 = a_1 + b_1 + c_1 + \cdots + n_1 \\ s_2 = a_2 + b_2 + c_2 + \cdots + n_2 \\ \quad \text{etc.} \qquad \text{etc.} \qquad \text{etc.} \end{array}\right\} \tag{9}$$

and form the quantities $[as]$ $[bs]\cdots[sn]$. It will appear from the mode in which the coefficients of the normal equations are formed that

Check

$$\left.\begin{array}{l} [aa]+[ab]+[ac]+\cdots+[an]=[as] \\ [ab]+[bb]+[bc]+\cdots+[bn]=[bs] \\ \qquad\quad \text{etc.} \qquad\quad \text{etc.} \qquad\quad \text{etc.} \end{array}\right\} \tag{9a}$$

The $[as]$, $[bs]$, etc., are formed in precisely the same manner as $[ab]$, $[ac]$, etc., and the check relations above given must be satisfied by the computed values of these quantities.

Where only two unknown quantities are involved in the normal equations the solution of the equations may be conveniently made by any of the methods of elementary algebra, but if the number of unknowns is greater than two, the simple and elegant method of successive substitutions proposed by *Gauss* may be employed with advantage.

The normal equations in the case of three unknown quantities are:

$$\left.\begin{array}{l} [aa]x+[ab]y+[ac]z+[an]=0 \\ [ab]x+[bb]y+[bc]z+[bn]=0 \\ [ac]x+[bc]y+[cc]z+[cn]=0 \end{array}\right\} \tag{10}$$

and from the first of these

$$x=-\frac{[ab]}{[aa]}y-\frac{[ac]}{[aa]}z-\frac{[an]}{[aa]}$$

This value of x substituted in the second and third equations transforms them into

$$\left.\begin{aligned}[bb\cdot 1]y+[bc\cdot 1]z+[bn\cdot 1]&=0\\ [bc\cdot 1]y+[cc\cdot 1]z+[cn\cdot 1]&=0\end{aligned}\right\} \tag{11}$$

in which

$$\left.\begin{aligned}[bb\cdot 1]&\equiv[bb]-\frac{[ab]}{[aa]}[ab] \quad [cc\cdot 1]\equiv[cc]-\frac{[ac]}{[aa]}[ac]\\ [bc\cdot 1]&\equiv[bc]-\frac{[ab]}{[aa]}[ac] \quad [bc\cdot 1]\equiv[bc]-\frac{[ac]}{[aa]}[ab]\\ [bn\cdot 1]&\equiv[bn]-\frac{[ab]}{[aa]}[an] \quad [cn\cdot 1]\equiv[cn]-\frac{[ac]}{[aa]}[an]\end{aligned}\right\} \tag{12}$$

These equations constitute a new set of normals, from which one unknown quantity has been eliminated. The correctness of the numerical work of this elimination may be tested by a continuation of the checks used in forming the original normals. We introduce an auxiliary quantity

$$[bs\cdot 1]=[bb\cdot 1]+[bc\cdot 1]+[bn\cdot 1]$$

and inquire its relation to $[as]$, $[bs]$, etc. If we substitute in the expression for $[bs\cdot 1]$ the values of $[bb\cdot 1]$, $[bc\cdot 1]$, $[bn\cdot 1]$ in terms of the original coefficients, having regard to the relations

$$\left.\begin{aligned}[aa]+[ab]+[ac]+[an]&=[as]\\ [ab]+[bb]+[bc]+[bn]&=[bs]\end{aligned}\right\} \tag{13}$$

$$\left.\begin{aligned}&\text{we find} \quad [bs\cdot 1]=[bs]-[ab]-\frac{[ab]}{[aa]}\{[as]-[aa]\}\\ &\text{whence} \quad [bs\cdot 1]=[bs]-\frac{[ab]}{[aa]}[as]\end{aligned}\right\} \tag{14}$$

and similarly,

$$[cs\cdot 1]=[cs]-\frac{[ac]}{[aa]}[as]$$

We may therefore obtain a complete check upon the accuracy of the numerical work involved in the elimination of x, by

forming the quantities $[bs \cdot 1]$, $[cs \cdot 1]$, in the same manner as $[bb \cdot 1]$, $[bc \cdot 1]$, $[bn \cdot 1]$, etc., and comparing the actual sums of these latter quantities with the computed check quantities.

By a repetition of the process of elimination we obtain

$$[cc \cdot 2]z + [cn \cdot 2] = 0 \qquad \textit{Check} \quad [cs \cdot 2]$$

where

$$\left.\begin{aligned} [cc \cdot 2] &= [cc \cdot 1] - \frac{[bc \cdot 1]}{[bb \cdot 1]}[bc \cdot 1] \\ [cn \cdot 2] &= [cn \cdot 1] - \frac{[bc \cdot 1]}{[bb \cdot 1]}[bn \cdot 1] \\ [cs \cdot 2] &= [cs \cdot 1] - \frac{[bc \cdot 1]}{[bb \cdot 1]}[bs \cdot 1] \\ &= [cc \cdot 2] + [cn \cdot 2] \end{aligned}\right\} \tag{15}$$

and we are enabled to write the following equivalents for the original normal equations.

ELIMINATION EQUATIONS.

$$\left.\begin{aligned} x + \frac{[ab]}{[aa]} \cdot y + \frac{[ac]}{[aa]} \cdot z + \frac{[an]}{[aa]} &= 0 \\ y + \frac{[bc \cdot 1]}{[bb \cdot 1]} \cdot z + \frac{[bn \cdot 1]}{[bb \cdot 1]} &= 0 \\ z + \frac{[cn \cdot 2]}{[cc \cdot 2]} &= 0 \end{aligned}\right\} \tag{16}$$

The last of these equations gives the value of z directly, the second furnishes y as soon as z is known, and the first gives the value of x. The whole solution is therefore reduced to finding the values of the coefficients and absolute terms in these *elimination equations*. A convenient arrangement of the computation by which these quantities are obtained is given in the following example, in which the actual computation is exhibited upon one page, and the opposite page contains a schedule correspondingly arranged showing the analytical equivalent of each number contained in the computation.

In making the multiplications of $[ab]$, $[ac]$, $[an]$, $[as]$, by the constant factor $\frac{[ab]}{[aa]}$, the logarithm of this factor is written on the edge of a slip of paper, and being held successively adjacent to the logarithms of $[ab]$, $[ac]$, $[an]$, $[as]$, the sum of the two logarithms is taken mentally, the corresponding number looked out from a logarithmic table and written in its proper place under $[bb]$, $[bc]$, $[bn]$, $[bs]$, a subtraction then gives the value of $[bb \cdot 1]$, $[bc \cdot 1]$, $[bn \cdot 1]$, $[bs \cdot 1]$, and a similar process is followed for each other derived coefficient.

§ **10. Example.** To illustrate the principles contained in the preceding sections, and to exhibit in detail the process of deriving the most probable values of several unknown quantities which are connected with the observed quantity by a rather complicated relation, we select from Vol. iii. Part 1 of the *Memoirs of the National Academy of Sciences*, page 58, the following series of experiments made with a 10-gauge Colt gun, loaded with uniform charges of four drams of powder and $1\frac{1}{4}$ ounces of shot, the shot ranging in fineness from No. 10 up to No. 1 Buck. The purpose of the experiments was to determine the relation existing between the size (fineness) of the shot and its average velocity over a range of 30 yards. The following table contains the results of the experiments, each velocity being the mean result of from three to six discharges of the gun. The weight of a pellet of No. 10 shot is taken as the unit of weight, and the velocities are expressed in feet per second.

Size.	Weight.	Observed Velocity.
No. 10	1	848
8	2	920
6	4	966
3	8	989
BB.	16	1000
FF.	32	1017
No. 1 Buck	64	1067

By plotting these results in a curve with the weight of the shot as abscissas, and the observed velocities as ordinates, the

experimenter reached the conclusion that the relation between the weight W and the velocity V, is expressed by an equation of the form

$$\frac{V}{l} - \sec^{-1}\frac{W^n}{m} = 0$$

in which l, m, n, are constants whose values are to be determined from the observations. It will be found upon trial that $l_0 = 700$, $m_0 = 0.28$, $n_0 = 0.42$, in connection with the observed values of V and W will approximately satisfy this equation, and we therefore adopt these approximate values and proceed (§ 8) to determine the corrections Δl, Δm, Δn, which when added to l_0, m_0, n_0, will furnish the most probable values of l, m, n.

The several differential coefficients of the observation equation, $f(l, m, n, V) = 0$ are

$$\frac{dV}{dl_0} = \frac{V}{l_0}$$

$$\frac{dV}{dm_0} = -\frac{l_0}{m_0}\cot\frac{V}{l_0}$$

$$\frac{dV}{dn_0} = \frac{l_0}{M}\log W \cot\frac{V}{l_0}$$

in which M denotes the modulus of the common system of logarithms, $M = 0.43429$. In the factor, $\cot\frac{V}{l_0}$, $\frac{V}{l_0}$ is the ratio of two numbers, and must be construed as representing a certain arc expressed in parts of the radius: the corresponding arc expressed in degrees is $57°.2958\frac{V}{l_0}$.

The form of the observation equation with which we are here concerned is

$$\frac{V}{l_0}\cdot\Delta l - \frac{l_0}{m_0}\cot\frac{V}{l_0}\cdot\Delta m + \frac{l_0}{M}\log W\cot\frac{V}{l_0}\Delta n + \left(V - l_0\sec^{-1}\frac{W^{n_0}}{m_0}\right) = 0$$

and introducing into this equation the numerical values of l_0, m_0, n_0, V, W, M, we find the following

OBSERVATION EQUATIONS.

(1)	$1.29\,\Delta l$	$-729\,\Delta m+$	$0\,\Delta n+$	$53=0$	$p=1.0$
(2)	1.36	-535	$+104$	$+32=0$	1.0
(3)	1.41	-396	$+154$	$+24=0$	0.8
(4)	1.45	-294	$+172$	$+28=0$	1.0
(5)	1.48	-219	$+170$	$+38=0$	1.2
(6)	1.50	-164	$+159$	$+37=0$	1.1
(7)	1.52	-122	$+142$	$-2=0$	1.0

The absolute terms of these equations are residuals obtained by substituting in the original equation

$$\frac{V}{l}-\sec^{-1}\frac{W^n}{m}=0$$

the assumed values of l_0, m_0, and n_0, and the smallness of these residuals compared with the values of V, shows that the assumed quantities are approximately correct values of l, m, n. The memoir from which our data are taken contains no indication of the weights to be assigned to the several determinations of V, and in the absence of such information they should all be treated as equally precise and given the weight 1; but for the sake of illustration a slightly different set of weights indicated above by p has been assigned to them, and by multiplying each equation by the square root of its weight we obtain the following

WEIGHTED OBSERVATION EQUATIONS.

$1.29\,\Delta l$	$-729\,\Delta m+$	$0\,\Delta n+$	$53=0$
1.36	-535	$+104$	$+32=0$
1.26	-352	$+137$	$+21=0$
1.45	-294	$+172$	$+28=0$
1.62	-239	$+185$	$+42=0$
1.58	-172	$+167$	$+38=0$
1.52	-122	$+142$	$-2=0$

The coefficients and absolute terms in these equations are of very different magnitudes, and to simplify the subsequent numerical work we divide each equation through by 100 and put

$$\Delta l = \frac{x}{0.0162}, \qquad \Delta m = \frac{y}{7.29}, \qquad \Delta n = \frac{z}{1.85}$$

and introduce x, y, z into the equations in place of Δl, Δm, Δn. This step, which is frequently called rendering the equations homogeneous, furnishes the following

HOMOGENEOUS WEIGHTED OBSERVATION EQUATIONS.

$0.796x$	$- 1.000y$	$+ 0.000z$	$+ 0.530 = 0$	$s = + 0.326$
0.839	$- 0.733$	$+ 0.563$	$+ 0.320 = 0$	0.989
0.777	$- 0.482$	$+ 0.741$	$+ 0.210 = 0$	1.246
0.895	$- 0.403$	$+ 0.931$	$+ 0.280 = 0$	1.703
1.000	$- 0.327$	$+ 1.000$	$+ 0.420 = 0$	2.093
0.975	$- 0.236$	$+ 0.903$	$+ 0.380 = 0$	2.022
0.938	$- 0.167$	$+ 0.768$	$- 0.020 = 0$	1.519

The values of $s = a + b + c + n$, which are to be used as a check in the formation of the normal equations, are derived from these equations.

The formation of the coefficients of the normal equations by the use of a table of squares, *Bessel's* method, is represented in the following tables:

SUMS OF THE COEFFICIENTS.

Equation	$a+b$	$a+c$	$a+n$	$a+s$	$b+c$	$b+n$	$b+s$	$c+n$	$c+s$
1....	0.204	0.796	1.326	1.122	1.000	0.470	0.674	0.530	0.326
2....	.106	1.402	1.159	1.828	.170	.413	0.256	0.883	1.552
3....	.295	1.518	0.987	2.023	.259	.272	0.764	0.951	1.987
4....	.492	1.826	1.175	2.598	.528	.123	1.300	1.211	2.634
5....	.673	2.000	1.420	3.093	.673	.093	1.766	1.420	3.093
6....	.739	1.878	1.355	2.997	.667	.144	1.786	1.283	2.925
7....	.771	1.706	0.918	2.457	.601	.187	1.352	0.748	2.287

SQUARES.

Equation.	a^2	$(a+b)^2$	$(a+c)^2$	$(a+n)^2$	$(a+s)^2$	b^2	$(b+c)^2$	$(b+n)^2$	$(b+s)^2$	c^2	$(c+n)^2$	$(c+s)^2$	n^2	s^2
1	0.634	0.042	0.634	1.758	1.259	1.000	1.000	0.221	0.454	0.000	0.281	0.106	0.281	0.106
2	.704	.011	1.966	1.343	3.342	.537	.029	.171	0.066	.317	0.780	2.409	.102	0.978
3	.604	.087	2.304	0.974	4.094	.232	.067	.074	0.584	.549	0.904	3.948	.044	1.552
4	.801	.242	3.334	1.381	6.750	.162	.279	.015	1.690	.867	1.466	6.938	.078	2.900
5	1.000	.453	4.000	2.016	9.567	.107	.453	.009	3.119	1.000	2.022	9.567	.176	4.381
6	.951	.546	3.527	1.836	8.982	.056	.445	.021	3.190	.815	1.646	8.556	.144	4.088
7	.880	.594	2.910	0.843	6.037	.028	.361	.035	1.828	.590	0.560	5.230	.000	2.307
	5.576	1.975	18.675	10.151	40.031	2.122	2.634	0.546	10.931	4.138	7.659	36.754	0.825	16.312
	. . .	7.698	9.714	6.401	21.888	. . .	6.260	2.947	18.434	. . .	4.963	20.450	. . .	. . . .
	+ 5.576	− 2.861	+ 4.480	+ 1.875	+ 9.071	+ 2.122	− 1.813	− 1.200	− 3.751	+ 4.138	+ 1.348	+ 8.152	+ 0.825	+ 16.312
	$[aa]$	$[ab]$	$[ac]$	$[an]$	$[as]$	$[bb]$	$[bc]$	$[bn]$	$[bs]$	$[cc]$	$[cn]$	$[cs]$	$[nn]$	$[ss]$
					+ 9.070				− 3.752			+ 8.153		

From the sums of the squares contained in the several columns of this table the coefficients $[ab]$, $[ac]$, etc., are computed at the foot of the columns by the relations

$$[ab] = \tfrac{1}{2}\{[(a+b)^2] - ([a^2] + [b^2])\}, \text{ etc.}$$

The check quantity $[as]$ is compared with

$$[aa] + [ab] + [ac] + [an]$$

whose value is written immediately under $[as]$, and which must agree with $[as]$ within two or three units of the last decimal place. Every coefficient of the normal equations enters into one or more of these sums, which therefore furnish a complete test of the accuracy of the work in passing from the homogeneous observation equations to the normal equations.

We now write the

NORMAL EQUATIONS.

$$+ 5.576x - 2.861y + 4.480z + 1.875 = 0$$
$$- 2.861x + 2.122y - 1.813z - 1.200 = 0$$
$$+ 4.480x - 1.813y + 4.138z + 1.348 = 0$$

It may be seen from an inspection of these equations that the data upon which they are based will not furnish a good determination of the values of all the unknowns, for if the first equation be divided by -2 the quotient will be very like the second equation, and if it be multiplied by $+\frac{4}{5}$ the product will be very like the third equation. We proceed, however, with the solution by *Gauss'* method, which will furnish the best results that the data can be made to yield.

Solution of the Normal Equations.

	$[aa]$	$[ab]$	$[ac]$	$[an]$	$[as]$
	$\log[aa]$	$\log[ab]$	$\log[ac]$	$\log[an]$	$\log[as]$
		$[bb]$	$[bc]$	$[bn]$	$[bs]$
$\log\frac{[ab]}{[aa]}$		$\frac{[ab]}{[aa]}[ab]$	$\frac{[ab]}{[aa]}[ac]$	$\frac{[ab]}{[aa]}[an]$	$\frac{[ab]}{[aa]}[as]$
		$[bb\cdot 1]$	$[bc\cdot 1]$	$[bn\cdot 1]$	$[bs\cdot 1]$
					Check sum.
		$\log[bb\cdot 1]$	$\log[bc\cdot 1]$	$\log[bn\cdot 1]$	$\log[bs\cdot 1]$
			$[cc]$	$[cn]$	$[cs]$
$\log\frac{[ac]}{[aa]}$			$\frac{[ac]}{[aa]}[ac]$	$\frac{[ac]}{[aa]}[an]$	$\frac{[ac]}{[aa]}[as]$
			$[cc\cdot 1]$	$[cn\cdot 1]$	$[cs\cdot 1]$
					Check sum.
$\log\frac{[bc\cdot 1]}{[bb\cdot 1]}$			$\frac{[bc\cdot 1]}{[bb\cdot 1]}[bc\cdot 1]$	$\frac{[bc\cdot 1]}{[bb\cdot 1]}[bn\cdot 1]$	$\frac{[bc\cdot 1]}{[bb\cdot 1]}[bs\cdot 1]$
			$[cc\cdot 2]$	$[cn\cdot 2]$	$[cs\cdot 2]$
					Check sum.
$\log\frac{[cn\cdot 2]}{[cc\cdot 2]}$			$\log[cc\cdot 2]$	$\log[cn\cdot 2]$	

Elimination Equations.

$$x + \frac{[ab]}{[aa]}y + \frac{[ac]}{[aa]}z + \frac{[an]}{[aa]} = 0$$

$$y + \frac{[bc\cdot 1]}{[bb\cdot 1]}z + \frac{[bn\cdot 1]}{[bb\cdot 1]} = 0$$

$$z + \frac{[cn\cdot 2]}{[cc\cdot 2]} = 0$$

The course of the computation after the formation of the elimination equations is sufficiently indicated upon the opposite page.

Solution of the Normal Equations.

	$+5.576$	-2.861	$+4.480$	$+1.875$	$+9.070$
	0.7464	$0.4566\,n$	0.6513	0.2730	0.9576
		$+2.122$	-1.813	-1.200	-3.752
$9.7102\,n$		$+1.468$	-2.299	-0.962	-4.654
		$+0.654$	$+0.486$	-0.238	$+0.902$
					$+0.902$
		9.8156	9.6866	$9.3766\,n$	9.9552
			$+4.138$	$+1.348$	$+8.153$
9.9049			$+3.599$	$+1.507$	$+7.286$
			$+0.539$	-0.159	$+0.867$
					$+0.866$
9.8710			$+0.361$	-0.177	$+0.670$
			$+0.178$	$+0.018$	$+0.197$
					$+0.196$
9.0049			9.2504	8.2553	...

Elimination Equations.

$$x - 0.513\,y + 0.803\,z + 0.336 = 0 \qquad x = -0.030$$
$$y + 0.743\,z - 0.364 = 0 \qquad y = +0.439$$
$$z + 0.101 = 0 \qquad z = -0.101$$

$\log x$	$8.4771\,n$	$\log y$	9.6425	$\log z$	$9.0049\,n$
$\log 0.0162$	8.2095	$\log 7.29$	0.8627	$\log 1.85$	0.2672
Δl	-1.8	Δm	$+0.0602$	Δn	-0.0547
l_0	700.0	m_0	$+0.2800$	n_0	$+0.4200$
l	698.2	m	$+0.3402$	n	$+0.3653$

If with the values of l, m, n thus obtained the corresponding velocities be computed by means of the original equation

$$\frac{V}{l} = \sec^{-1}\frac{W^n}{m}$$

the resulting residuals should be smaller than those derived from the substitution of l_0, m_0, n_0, i.e., the absolute terms of the observation equations. The following comparison of these residuals shows a much better representation of the observed values of V, especially if the sums of the squares, $[vv]$, be compared.

	Observed — Computed V.						
Weight of Shot	1	2	4	8	16	32	64
$f(l_0\, m_0\, n_0)$	− 53,	− 32,	− 24,	− 28,	− 38,	− 37,	+ 2
$f(l,\ m,\ n)$	− 6,	+ 10,	+ 13,	+ 4,	− 10,	− 13,	+ 22

Not only are the residuals diminished in magnitude, but their distribution is much more nearly in agreement with the law of error.

The values thus obtained for l, m, n ought not to be considered the best attainable, since the corrections Δm, Δn are relatively large fractions of m_0 and n_0, and it is probable that the neglected terms containing $\overline{\Delta m}^2$, $\overline{\Delta n}^2$, etc., have an appreciable influence upon the solution. To secure the utmost accuracy these values of l, m, n should be treated as new approximations and another set of corrections Δl, Δm, Δn derived. This resolution is recommended to the student as a valuable exercise. Let the student also derive from the data of § 1 the most probable values of l_0 and c, assigning unequal weights to the several equations.

§ **11. Conditioned Observations.** There is a class of cases in which the application of the principle of least squares seems to produce absurd results. Thus if each angle of a plane triangle be measured many times in order to obtain an accurate set of values for the angles, the application of the principle that the $[pvv]$ must be made a minimum will furnish as the most probable value of each angle the weighted mean of the

measures of that angle, but the sum of these weighted means will usually differ slightly from 180°, and since the sum of the angles of every plane triangle must equal 180° it appears that the most probable values above derived are impossible values. It must, however, be noted that the method of treatment above outlined is itself a violation of Principle A, § 1, since the knowledge that the sum of the angles must equal 180° furnishes a relation among those angles which may be used and ought to be used in determining their most probable values; and the apparent absurdity above found is produced by neglecting this part of the data.

A relation such as the above which must be exactly satisfied by a set of observed quantities is called a rigorous condition, the equation by which the relation is expressed is called an equation of condition, and observations of such quantities are known as conditioned observations. The number of rigorous conditions is, of course, always less than the number of unknown quantities, since if it were equal to the number of such quantities the values of the latter would be determined by the conditions alone, independently of any observations.

In order to develop a convenient method of treating rigorous conditions, let x, y, z be three unknown quantities which are to be determined from observation, but whose values are required to satisfy the equations of condition

$$\phi(x, y, z) = 0 \qquad \psi(x, y, z) = 0$$

Let the measurements or observations for the determination of the unknown quantities be represented by observation equations of the form

$$f_1(x, m) = 0 \qquad f_2(y, n) = 0 \qquad f_3(z, q) = 0$$

m, n, and q being the quantities directly measured, and the measures for the determination of x being quite independent of those for y, z, etc. In accordance with the principles of least squares the values of the unknown quantities are to be so determined that $[pvv]$ shall be made a minimum in each series of observations above represented, and therefore the sum of all

the weighted squares of the residuals must also be a minimum. Owing to the conditions $\phi(x, y, z) = 0$, $\psi(x, y, z) = 0$ it will not in general be possible to assign to the unknown quantities values which will give to $[pvv]$ its least possible value, and the problem becomes one of conditioned or relative minima, i.e. out of all the sets of values of x, y, z which will exactly satisfy the equations of condition it is required to find that set which assigns to $[pvv]$ its least value consistent with those equations.

The method of determining relative minima is as follows: (*Jordan, Cours d'Analyse*, Vol. i., § 205). Multiply each equation of condition by an undetermined constant factor, and add the products to the function which is to be made a minimum. The derivative of the new function with respect to each unknown quantity must be placed equal to 0, and the equations thus formed, together with the equations of condition, will be just sufficient to determine the unknown quantities and the constant multipliers. Thus, in the present case, representing the multipliers by $-2k_1$ and $-2k_2$, we have for the new function

$$w = [pvv] - 2k_1\phi(x, y, z) - 2k_2\psi(x, y, z) \qquad (17)$$

and

$$\frac{dw}{dx} = 0 \qquad \frac{dw}{dy} = 0 \qquad \frac{dw}{dz} = 0 \qquad (18)$$

$$\phi(x, y, z) = 0 \qquad \psi(x, y, z) = 0$$

will determine k_1, k_2, x, y, and z.

It was shown in § 7 that in general for three unknown quantities,

$$\frac{d[pvv]}{dx} = 2[paa]x + 2[pab]y + 2[pac]z + 2[pan]$$

but in the case here considered those observation equations which contain x do not contain either y or z, and, therefore, the b and c coefficients in those equations are to be considered zero, all the products ab, ac, are also zero, and

$$\frac{d[pvv]}{dx} = 2[paa]x + 2[pan]$$

with similar expressions for the y and z derivatives.

Denoting for the sake of brevity $\phi(x, y, z)$ and $\psi(x, y, z)$ by ϕ and ψ respectively, we obtain by differentiating w

$$\tfrac{1}{2}\frac{dw}{dx} = [paa]x + [pan] - k_1\frac{d\phi}{dx} - k_2\frac{d\psi}{dx} = 0$$

$$\tfrac{1}{2}\frac{dw}{dy} = [pbb]y + [pbn] - k_1\frac{d\phi}{dy} - k_2\frac{d\psi}{dy} = 0$$

$$\tfrac{1}{2}\frac{dw}{dz} = [pcc]z + [pcn] - k_1\frac{d\phi}{dz} - k_2\frac{d\psi}{dz} = 0$$

from which

$$x = -\frac{[pan]}{[paa]} + \frac{d\phi}{dx}\cdot\frac{k_1}{[paa]} + \frac{d\psi}{dx}\cdot\frac{k_2}{[paa]}$$

$$y = -\frac{[pbn]}{[pbb]} + \frac{d\phi}{dy}\cdot\frac{k_1}{[pbb]} + \frac{d\psi}{dy}\cdot\frac{k_2}{[pbb]}$$

$$z = -\frac{[pcn]}{[pcc]} + \frac{d\phi}{dz}\cdot\frac{k_1}{[pcc]} + \frac{d\psi}{dz}\cdot\frac{k_2}{[pcc]}$$

These equations determine the values of x, y, z, when k_1 and k_2 are known, and it should be observed that the first terms of the second members of the equations are the values of x, y, z, which would be obtained by treating the observations as if these quantities were entirely independent of each other, e.g. in the case of direct observations of the quantities they are the weighted means of the observations. If we represent the values thus obtained by x_0, y_0, z_0, and represent by v_1, v_2, v_3 the corrections which must be added to these quantities in order to obtain the most probable values of x, y, z, i.e. put

$$x = x_0 + v_1 \qquad y = y_0 + v_2 \qquad z = z_0 + v_3$$

we shall have

$$\left.\begin{aligned} v_1 &= \frac{d\phi}{dx}\cdot\frac{k_1}{[paa]} + \frac{d\psi}{dx}\cdot\frac{k_2}{[paa]} \\ v_2 &= \frac{d\phi}{dy}\cdot\frac{k_1}{[pbb]} + \frac{d\psi}{dy}\cdot\frac{k_2}{[pbb]} \\ v_3 &= \frac{d\phi}{dz}\cdot\frac{k_1}{[pcc]} + \frac{d\psi}{dz}\cdot\frac{k_2}{[pcc]} \end{aligned}\right\} \qquad (19)$$

The quantities k_1 and k_2 are called *correlates*, and from the manner in which they were introduced it appears that the number of correlates is equal to the number of rigorous conditions to which the observed quantities are subject. To determine the values of the correlates let x_0+v_1, y_0+v_2, z_0+v_3 be substituted for x, y, z in the equations of condition, and the equations developed by Taylor's Formula, giving for the $\phi(x, y, z)$

$$\phi(x_0, y_0, z_0)+\frac{d\phi}{dx}v_1+\frac{d\phi}{dy}v_2+\frac{d\phi}{dz}v_3+\text{etc.}=0$$

and a similar expression for $\psi(x, y, z)$. Let the values of v_1, v_2, v_3 in terms of k_1 and k_2 be substituted in these equations, and put

$$\left.\begin{aligned} \frac{\frac{d\phi}{dx}}{\sqrt{[paa]}}&=a_1, & \frac{\frac{d\phi}{dy}}{\sqrt{[pbb]}}&=a_2, & \frac{\frac{d\phi}{dz}}{\sqrt{[pcc]}}&=a_3,\\ \frac{\frac{d\psi}{dx}}{\sqrt{[paa]}}&=\beta_1, & \frac{\frac{d\psi}{dy}}{\sqrt{[pbb]}}&=\beta_2, & \frac{\frac{d\psi}{dz}}{\sqrt{[pcc]}}&=\beta_3, \end{aligned}\right\} \qquad (20)$$

and the equations become

$$\left.\begin{aligned} [aa]k_1+[a\beta]k_2+\phi(x_0, y_0, z_0)&=0\\ [a\beta]k_1+[\beta\beta]k_2+\psi(x_0, y_0, z_0)&=0 \end{aligned}\right\} \qquad (21)$$

from which the values of k_1 and k_2 may be obtained, and thus the values of v_1, v_2, v_3 from equations (19).

The method by which the above equations have been derived for the case of three unknown quantities connected by two equations of condition is perfectly general and may be extended to any other number of quantities whose values are to be obtained from independent observations. In the cases which actually arise in practice the observation equations and equations of condition are usually of simple form, the differential coefficients and the quantities a, b, c, etc., being usually equal to either 1 or 0.

PROBLEM. Let the student show by the method of correlates that if the sum of the measured angles of a plane triangle exceed 180° by a quantity e, the angles must be corrected by distributing e among them in such a manner that the correction to each angle is inversely proportional to the weight of the angle.

To illustrate the application of the principles of the present section to a numerical problem, we select from the *U. S. C. & G. Survey Report* for 1884, pages 409 *et seq.*, the following telegraphic determinations of longitude, and seek to adjust them so that they shall be mutually consistent. Each difference of longitude between two stations was directly observed, so that the observation equations are all of the form $x = m_1$, $x = m_2$, etc., and the values given below are the weighted means of the individual observations of each series. The probable error of each determination (see § 12) is placed immediately after the quantity itself, and the weights of the determinations are assumed to be inversely proportional to the squares of the probable errors.

Stations.	Symbol.	Observed Difference of Longitude.	$\frac{1}{\sqrt{p}}$	$\frac{1}{p}$
Cambridge, Mass., Washington, D.C.,	x_0	$23^m\ 41^s.041 \pm 0^s.018$	0.18	0.032
Cambridge, Mass., Cleveland, O.,	y_0	$42\ \ 14.875 \pm 0.038$	0.38	.144
Cambridge, Mass., Columbus, O.,	z_0	$47\ \ 27.713 \pm 0.035$	0.35	.122
Washington, D.C., Columbus, O.,	u_0	$23\ \ 46.816 \pm 0.038$	0.38	.144
Cleveland, O., Columbus, O.,	w_0	$5\ \ 12.929 \pm 0.045$	0.45	.202

The five observed differences of longitude give rise to two rigorous conditions represented by the following equations of condition:

$$\phi(\,),\ u + x - z = 0 \qquad \psi(\,),\ w + y - z = 0$$

The coefficients in the observation equations being all equal to unity,

$$[paa] = p_x, \quad [pbb] = p_y, \text{ etc.},$$

and $\alpha_1 = \frac{d\phi}{dx} \cdot \frac{1}{\sqrt{p_x}}$, $\alpha_2 = \frac{d\phi}{dy} \cdot \frac{1}{\sqrt{p_y}}$, $\beta_1 = \frac{d\psi}{dx} \cdot \frac{1}{\sqrt{p_x}}$, etc.,

and from these expressions are derived the following values of the coefficients, together with the sums $s_1 = \alpha_1 + \beta_1$, $s_2 = \alpha_2 + \beta_2$, etc., which are to be employed as a check upon the formation of the normal equations for determining the correlates.

Coefficients.

Subscripts.	1.	2.	3.	4.	5.
α	+0.18	0.00	−0.35	+0.38	0.00
β	0.00	+0.38	−0.35	0.00	+0.45
s	+0.18	+0.38	−0.70	+0.38	+0.45

Formation of the Correlate Equations.

$\alpha\alpha$	$\alpha\beta$	αs	$\beta\beta$	βs
+0.0324	+0.0000	+0.0324	+0.0000	0.0000
.0000	.0000	.0000	.1444	.1444
.1225	.1225	.2450	.1225	.2450
.1444	.0000	.1444	.0000	.0000
.0000	.0000	.0000	.2025	.2025
+0.2993	+0.1225	+0.4218	+0.4694	+0.5919
Check.		0.4218		0.5919

Correlate Normal Equations.

$$+0.2993\,k_1 + 0.1225\,k_2 + 0^s.144 = 0 \qquad k_1 = -0^s.449$$
$$+0.1225\,k_1 + 0.4694\,k_2 + 0\,.091 = 0 \qquad k_2 = -0\,.078$$

The absolute terms of the correlate equations are obtained by substituting the observed values x_0, y_0, z_0, u_0, w_0 in the equations of condition, and the values of k_1, k_2 may be found from the correlate equations, either by Gauss's method of substitution or by any of the ordinary algebraic processes of elimination. The corrections to x_0, y_0, z_0, etc., and the adopted values of the unknown quantities, are now found from

$$v_1 = +0.032\,k_1 + 0.000\,k_2 = -0^s.014 \qquad x = 23^m\,41^s.027$$
$$v_2 = +0.000\,k_1 + 0.144\,k_2 = -0.011 \qquad y = 42\ \ 14.864$$
$$v_3 = -0.122\,k_1 - 0.122\,k_2 = +0.064 \qquad z = 47\ \ 27.777$$
$$v_4 = +0.144\,k_1 + 0.000\,k_2 = -0.065 \qquad u = 23\ \ 46.751$$
$$v_5 = +0.000\,k_1 + 0.202\,k_2 = -0.016 \qquad w = \ \ 5\ \ 12.913$$

The values thus obtained satisfy the rigorous conditions of the problem, and are the most probable values which can be obtained from the data given above.

§ **12. The Probable Error.** Every intelligent observer desires to know something of the quality of his observations, how good or how bad they are; the computer who has to combine the results of different series of observations should have some knowledge of their relative accuracy in order to assign to each series its proper weight; and the investigator engaged in a complicated series of experiments desires some criterion by which to estimate the relative errors of the several parts of his work, in order to properly apportion his care among them, giving the maximum attention where the greatest errors are to be feared. It is evident from the nature of the case that no absolute criterion of this kind can be furnished, since any series of observations may be affected with systematic errors which seriously impair the accuracy of its results but furnish no indication of their presence. Both observer and computer do, however, estimate the accuracy of observations by their agreement among themselves, and that within certain limits this procedure is correct follows from *Gauss's* law of error. If we suppose a very long series of observations affected only by accidental errors, the values of the unknown

quantities obtained from the series will differ but little from the true values (if the series is infinitely long they will be the true values), the residuals which they furnish will be very nearly the errors of observation, and the value of h in the equation of the error curve will furnish a measure of the precision of the observations as well as a measure of the smallness of the residuals. On the other hand, if the student attempts to construct the error curve corresponding to any short series of residuals, e.g., those of § 10, he will find that while they give him some information in regard to the curve there will be much that is arbitrary in its actual construction, and that many curves can be drawn which will appear to fit the residuals equally well, i.e. the amount of data in this case is insufficient to determine more than a rough approximation to the measure of precision of the observations. If the observations are affected with systematic errors, the residuals may be very different from the errors of the observations, and will then furnish no indication of their accuracy.

It thus appears that any conclusions in regard to the accuracy of a given set of observations must be treated with caution if they are based solely on the residuals furnished by the observations. Such conclusions are, in fact, valid only within certain limits whose general nature is indicated above, but within these limits the information thus furnished may be of much value, and it is frequently employed for the purposes indicated at the beginning of this section.

The measure of precision, h, seems to be indicated by its name as the appropriate means of expressing the average accuracy of a set of observations, but in practice it is not so used, another function of the residuals being found more convenient. If in a very long series of observations the residuals be arranged in the order of their numerical magnitude (without regard to sign), that residual which occupies the middle place in the series will have as many residuals greater than it as there are less than it, and in any future series of observations of the same degree of precision as that here considered, it will be an even chance that any given residual will be

greater than, or less than, the middle one above selected. This middle residual is usually denoted by r, and is rather inappropriately called the *probable error* of the series, the adjective having reference to the equal probabilities of the occurrence of residuals (errors) greater than, or less than, r.

It is apparent that the greater the precision of any set of observation, the smaller will be the corresponding probable error, but the exact relation which exists between h and r must be derived from the equation of the error curve. The symmetry of this curve with respect to the axis of y shows that the same law of distribution holds for both positive and negative errors, and that in a very long series of residuals the probable error r will occupy the middle place among the positive errors and among the negative errors considered separately, as well as among all the errors taken without regard to sign. Since we are concerned only with the numerical magnitude of r we may confine our attention to the positive residuals, and find the relation between r and h from that half of the error curve which lies to the right of the axis of y.

Since the probable error is a residual, it must be represented by the abscissa of some point on the axis of x, and we may determine this point from the condition that the ordinate drawn through it bisects the area of that half of the curve under consideration, since (from the relation between areas and the number of residuals of a given magnitude developed in § 4) this is the geometrical equivalent of the statement that the number of residuals greater than r is equal to the number less than r. By interpolation from the table in § 4, the value of the argument corresponding to $A = 0.25$ is found to be $hx = hr = 0.477$, whence the relation between the probable error and the measure of precision is

$$r = \frac{0.477}{h} \tag{22}$$

The student will observe that in the definition of the probable error reference is made to a very long series of observations, and in a series of infinite length the value of r might be found

REESE LIBRARY OF THE

immediately from its definition, but in any ordinary set of observations it is better to assume that the residuals are distributed in accordance with the law of error, and to determine the value of r from the relation between h and the sum of the squares of the residuals, § 5, which gives

$$r = \pm 0.477\sqrt{2}\sqrt{\frac{[vv]}{n}}$$

We here encounter a difficulty arising from the attempt to apply to a short series of residuals principles which are rigorously true only when the series is of infinite length. Suppose the above expression for r applied to a series of three observations involving three unknown quantities whose values are derived from the resulting observation equations. These values will exactly satisfy the equations, no matter what the errors of the observations may be, and the residuals being all zero, there will be found $r = 0$ and $h = \infty$, which is absurd. The observations in this case furnish no data from which to estimate their precision, and in every such case where the number of observations is equal to the number of unknown quantities, the expression for the probable error ought to become indeterminate, $\frac{0}{0}$. It is therefore customary to put

$$r = \pm 0.674\sqrt{\frac{[vv]}{n-\mu}} \tag{23}$$

in which μ denotes the number of quantities whose values have been derived from the observations. This equation, which is known as Bessel's expression for the probable error of a single observation, being only an approximate one, we may usually put $\frac{2}{3}$ in place of the coefficient 0.674. Among German physicists and astronomers, it is quite customary to suppress this coefficient altogether, and to use the "mean error"

$$\epsilon = \pm\sqrt{\frac{[vv]}{n-\mu}} \tag{24}$$

for the comparison of observations. Geometrically considered, ϵ denotes the abscissa of the point of inflexion of the error curve.

A simpler expression for the probable error may be obtained by substituting in the equation

$$r = \frac{0.477}{h}$$

a value of h derived as follows: Let each member of the equation of the error curve be multiplied by xdx and integrated between the limits $-\infty$ and $+\infty$, giving

$$\int_{-\infty}^{+\infty} xydx = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} xe^{-h^2x^2} dx$$

The value of the first integral in this equation is obviously 0, since as we pass along the error curve from $-\infty$ to $+\infty$ every value of y occurs once associated with a negative value of x, and again with a numerically equal positive value, and for every negative element $xydx$ in the integral there occurs an equal positive $xydx$ so that the entire sum is 0. If, however, we agree to neglect the sign of x and to condsider only its numerical value, we shall find

$$\int_{-\infty}^{+\infty} xydx = 2\int_{0}^{+\infty} xydx$$

and by a course of reasoning precisely similar to that applied in § 5 to the quantity $\int_{-\infty}^{+\infty} x^2ydx$, it may be shown that $2\int_{0}^{+\infty} xydx$ is equal to the mean of all the residuals taken without regard to sign. We may therefore write

$$\frac{[+v]}{n} = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} xe^{-h^2x^2} dx$$

where the + inside the brackets denotes that all of the residuals are to be treated as positive quantities. Putting $hx = t$ in the second member of this equation and remembering that here also we are concerned only with numerical values of x without regard to sign, we obtain

$$\frac{[+v]}{n} = \frac{2}{h\sqrt{\pi}} \int_{0}^{+\infty} te^{-t^2} dt = \frac{2}{h\sqrt{\pi}} \left(-\frac{e^{-t^2}}{2} \right)_{0}^{+\infty}$$

Introducing the limits into the integrated expression there results

$$\frac{[+v]}{n} = \frac{1}{h\sqrt{\pi}}$$

and $$r = 0.477\sqrt{\pi}\frac{[+v]}{n} \qquad (25)$$

This formula is rigorously correct only when the number of observations is infinite, and it must be transformed so as to become indeterminate when the number of observation equations is just sufficient to determine the unknown quantities, i.e., when $n = \mu$. This might be accomplished by writing $n - \mu$ in place of n, as was done in equation (19), but it is customary to substitute in this case $\sqrt{n(n-\mu)}$, which also renders r indeterminate when $n = \mu$, and gives values of r more nearly in agreement with equation (23). Making this substitution, we have

$$r = \pm 0.845\frac{[+v]}{\sqrt{n(n-\mu)}} \qquad (26)$$

which is known as Peters' formula for probable errors. This formula is very convenient for the numerical computation of probable errors, but where the number of observations is small the results furnished by equation (23) are considered more reliable, but neither formula can furnish a good determination of probable errors from a small number of observations.

The numerical application of these formulæ may be illustrated by the following short series of sextant observations for the determination of latitude.

Observations	v	vv		
43° 4′ 46″	19″	361	$\log[+v]$	2.403
4 24	3	9	*a.c.*, $\log\sqrt{n(n-1)}$	8.940 − 10
4 7	20	400	$\log 0.845$	9.926 − 10
4 28	1	1	$\log r$	1.269
4 59	32	1024	r	± 18″.6
4 39	12	144		
4 52	25	625	$\log[vv]$	3.886
4 52	25	625	$\log(n-1)$	1.041
3 47	40	1600	$\log\frac{[vv]}{n-1}$	2.845
4 15	12	144	$\log\sqrt{\frac{[vv]}{n-1}}$	1.422
3 36	51	2601	$\log 0.674$	9.829 − 10
4 40	13	169	$\log r$	1.251
Mean = 43 4 27	$[+v] = 253$	$[vv] = 7703$	r	± 17″.8
$n = 12$	$\mu = 1$			

The difference between the values of r found from the first and second powers of the residuals is small compared with the uncertainty of each arising from the small number of observations.

In so far as these observations can be considered as furnishing a value of r, they indicate that in a future series of similar and equally precise observations, there should be as many observations furnishing residuals (errors) greater than 18″ as there are observations giving residuals less than 18″. The ± which is commonly prefixed to the numerical value of r, denotes that the observed quantity is as apt to err in excess as in defect.

Let the student derive from the residuals given in § 10 a determination of the probable error of an observed V, noting that in this case $\mu = 3$.

§ 13. Probable Error of a Function of Observed Quantities. Let x', x'', x''' denote quantities which have been determined from observation, and let r', r'', r''' be their probable errors. Let u be a quantity whose value has been computed from the values of x', x'', x''' by means of the relation

$$f(x', x'', x''', u) = 0$$

It is evident that the precision with which u is determined, depends upon the precision of x', x'', x''', and by a slight extension of the term "probable error" we may consider the precision of u to be represented by a probable error, r, and may inquire the relation of r to r', r'', r'''.

Since a probable error is one of the residuals or errors of a very long series, we may obtain the desired relation between r, r', r'', r''' from a consideration of the general relation of any set of errors v', v'', v''', in x', x'', x''', to the corresponding error, v, in u. This relation is

$$v = \frac{df}{dx'}v' + \frac{df}{dx''}v'' + \frac{df}{dx'''}v''' + \text{etc.}$$

(see § 8). To avoid the necessity for considering the signs of v', v'', v''', let this equation be squared, giving

$$v^2 = \left(\frac{df}{dx'}\right)^2 v'^2 + \left(\frac{df}{dx''}\right)^2 v''^2 + \left(\frac{df}{dx'''}\right)^2 v'''^2$$

from which all terms involving the products $v'v''$, $v'v'''$, $v''v'''$, etc., have been dropped for the reason that the probable error of u depends upon the average magnitude of v, and in the long run any pair of residuals v', v'' will have opposite signs as often as they have like signs, and will therefore produce an equal number of positive and negative terms whose effect upon the mean value of v^2 will be very small compared with the terms containing v'^2, v''^2, v'''^2, which are always positive. Replacing these actual errors by the corresponding probable errors, we obtain

$$r^2 = \left(\frac{df}{dx'}\right)^2 r'^2 + \left(\frac{df}{dx''}\right)^2 r''^2 + \left(\frac{df}{dx'''}\right)^2 r'''^2 \qquad (27)$$

and an equation of similar form will express the relation of the probable error of the function to the probable errors of the quantities upon which it depends, whatever the number of these quantities may be.

We proceed to apply this relation to a few simple cases of frequent occurrence in practice.

(a) The probable error of the sum of n observed quantities.

In this case $u = x' + x'' + x''' + \cdots + x^n$

and each of the differential coefficients $\frac{df}{dx'}$, $\frac{df}{dx''}$, etc., equals 1;

whence $$r^2 = r'^2 + r''^2 + r'''^2 + \cdots + r^{n2} \qquad (28)$$

(b) The probable error of the mean of n observed quantities.

In this case $u = \frac{1}{n}(x' + x'' + x''' + \cdots + x^n)$

$$\frac{df}{dx'} = \frac{df}{dx''} = \text{etc.} = \frac{1}{n}$$

$$r^2 = \frac{1}{n^2}(r'^2 + r''^2 + r'''^2 + \cdots r^{n2})$$

We have here to distinguish two cases. If the x's are all of equal precision, the r's are equal, and may be represented by a common symbol r_1; whence

$$r = \pm\sqrt{\frac{nr_1^2}{n^2}} = \pm\frac{r_1}{\sqrt{n}} \tag{29}$$

If the observations are of unequal precision represented by weights p', p'', $p'''\cdots p^n$

we have

$$u = \frac{p'x' + p''x'' + p'''x''' + \cdots p^n x^n}{[p]}$$

$$\frac{df}{dx'} = \frac{p'}{[p]} \qquad \frac{df}{dx''} = \frac{p''}{[p]} \qquad \text{etc.}$$

$$r^2 = \frac{[p]}{[p]^2} r_1^2$$

$$r = \pm\frac{r_1}{\sqrt{[p]}} \tag{30}$$

where r_1 denotes the probable error of an observation whose weight is 1.

The relations here derived between the probable error of a single determination of a quantity, and the probable error of the mean of n determinations, may be employed in connection with equation (23), to determine the probable error of an adopted value based upon several determinations of a quantity. Thus, in the general case of observations of unequal weight, if r_1 represent the probable error of an observation of weight 1, and r the probable error of the weighted mean, we have from equation (23),

$$\left.\begin{aligned} r_1 &= 0.674\sqrt{\frac{[pvv]}{n-1}} \\ r &= \frac{r_1}{\sqrt{[p]}} \end{aligned}\right\} \tag{31}$$

Let the student show that when the observations are of equal precision and

$$u = a(x' - x'' - x''') \qquad r = \pm a\sqrt{3}r'$$

$$u = \sin\frac{x'}{a} \qquad r = \pm \cos\frac{x'}{a} \cdot \frac{r'}{a}$$

$$u = \log\left(\frac{x'}{x''}\right) \qquad r = \pm 0.434 \left\{\frac{1}{x'^2} + \frac{1}{x''^2}\right\}^{\frac{1}{2}} r'$$

§ 14. Assignment of Weights. Rejection of Observations. The term weight has been employed in the preceding sections as a measure of the quality of an observation, but its use is by no means limited to the case of single observations. Thus, from an *Investigation of the Distance of the Sun, etc.*, by *S. Newcomb,* we select the following determinations of the solar parallax.

Method by which determined.	Parallax.	Weight.
Meridian observations of Mars, 1862.	8″.855	25
Micrometric observations of Mars, 1862.	8.842	6
Parallactic inequality of the Moon.	8.838	16
Lunar equation of the Earth.	8.809	3
Transit of Venus, 1769.	8.860	6

Each value of the parallax here given is the final result of an elaborate discussion of many observations, and the weights indicate the relative excellence attributed to these results by the author of the investigation. If π denote any one of these values of the parallax, p its weight, and π_0 the most probable value of the parallax, we shall have

$$\pi_0 = \frac{[p\pi]}{[p]} = 8''.847 \qquad (\S\ 6)$$

It is to be noted that this value depends upon the weights assigned to the individual determinations, and that by properly selecting the weights, π_0 may be made to assume any value whatever between the least and the greatest single determination. Thus if the weight 100 be assigned to the value 8″.860, and to each of the other values the weight 1, we shall find $\pi_0 = 8''.859$, while a weight 100 for the value 8″.809 with a weight 1 for each of the others, makes $\pi_0 = 8''.811$. Between these limits the value of π_0 depends upon the judgment of the

computer in assigning weights, and this determination of weights is one of the most delicate questions that arise in the application of the method of least squares.

A relation between weights and probable errors may easily be established, which is frequently of service in that it enables the problem of weights to be stated in a different form. Let x denote an observation whose probable error is r and whose weight is 1, and denote by x', x'', x''', etc., observations or combinations of observations of the same quantity, whose weights and probable errors are represented by p', p'', p''', r', r'', r''', etc. In accordance with the definition of weights, x' is the equivalent of p' observations of the same quality as x, and from the equations derived in the preceding section we have

$$r' = \frac{r}{\sqrt{p'}}$$

with a similar expression for each of the other quantities; whence

$$p'r'^2 = p''r''^2 = p'''r'''^2 = r^2$$

and

$$p' = \frac{r^2}{r'^2}, \quad p'' = \frac{r^2}{r''^2}, \quad p''' = \frac{r^2}{r'''^2} \tag{32}$$

and, in general, the weights are inversely proportional to the squares of the probable errors.

It has been sufficiently shown that probable errors derived from the residuals furnished by a series of observations represent only the effects of accidental errors of observation, but we may extend the significance of the term so as to include an estimate of the effect upon x', x'', x''' of systematic errors in the observations. Let r_1 and r_2 represent those parts of the probable error which come from these two sources respectively, and from § 13 we find for their combined effect

$$r_0^2 = r_1^2 + r_2^2$$

and the expression for the weights becomes

$$p = \frac{r^2}{r_0^2} = \frac{r^2}{r_1^2 + r_2^2} \tag{33}$$

By this device the determination of weights is reduced to an estimate of the combined effect of accidental and systematic errors of observation upon the quantity whose weight is desired, and it was from an estimate of this character that the weights of the parallaxes given above were derived.

If r' denote the probable accidental error of a single observation, and the quantity whose weight is p is the mean of n such observations, we shall have

$$p = \frac{1}{\frac{r'^2}{n} + r_2^2} \tag{34}$$

from which the constant multiplier r^2 has been dropped, since only relative values of p are ever required. It appears from this equation that if the systematic errors, r_2, are very small compared with the accidental errors, r', the weight increases rapidly as the number of observations is increased, but if the systematic errors are large, the weight is but little affected by the number of observations; a relation to be considered in deciding how many observations shall be made to determine an unknown quantity.

In some cases it may be impossible to form any reliable estimate of the effect of systematic errors, and results which have been derived by different methods, or under different circumstances, may then be given equal weights on the supposition that they are affected by different systematic errors which it is equally important to eliminate; but this is equivalent to putting $r_2 = \infty$ in the equation for the weights, and it will rarely happen that this is the best estimate which can be made for the amount of the systematic errors.

It frequently happens that in a series of otherwise accordant observations, one or two will be found which differ widely from the others, and which if included in the final result, will furnish large residuals. What shall be done with observations of this kind has long been a vexed question. To reject them is equivalent to assigning to them the weight 0, and is the expression of the computer's judgment that they can contribute

nothing to the accuracy of the result which he seeks to obtain. In an infinitely long series of observations, errors of any finite magnitude may be permitted without impairing the accuracy of the final result, and the existence of such errors seems contemplated by the theory which we have adopted, since the equation of the error curve gives finite values of y for all values of x between the limits $-\infty$ and $+\infty$. But in the actual case which arises in practice where a result must be obtained from a comparatively small number of observations, a single one of these, if affected with a large error, may make the final result farther from the truth than any one of the other observations. On the other hand, cases are by no means unknown in which a single discordant observation in a series proves to be nearest to the true value of the quantity sought, the others having all been vitiated by some common cause; and between these extremes an infinite variety of cases may be found. It must in general remain a matter of doubt whether a given discordant observation should or should not be rejected, and the decision made by the computer must be his judgment based upon all the data available as to whether more will be gained by rejecting than by retaining it. A knowledge of the way in which observations are made, of the circumstances attending the particular observation in question, the magnitude of the errors which may reasonably be expected with the given observer and apparatus, or instrument, are elements which should be included in this judgment; and the observer will greatly facilitate its formation by making copious notes at the time of observation of all circumstances which in his opinion may affect the quality of his work, and particularly by noting any abnormal circumstances affecting a single observation or a part of the observations.

A doubtful observation should be rejected if it is the computer's deliberate judgment that its retention will hurt more than it will help his final result, but it is never legitimate for the computer to suppress an observation. A rejected observation should be included in the statement of his data, and may properly be accompanied by an explanation of the reasons for

its rejection, in order that any person interested in the result may form his own judgment of the data and the manner in which they have been discussed, and may, if necessary, rediscuss the observations in accordance with that judgment.

The conclusion of the whole matter of assigning weights to numerical data may be summed up in the statement that no mathematical expression will suffice for this purpose, but the weights must be determined by an exercise of personal judgment, and the wider the knowledge upon which this judgment is based, the greater confidence will the weights and the resulting values of the unknown quantities command.

§ **15. Empirical or Interpolation Formulæ.** In the preceding sections attention has been directed to that class of problems in which the theoretical relation between the observed quantities and those whose values are to be determined is known; that is, an equation of known form exists between them, and the problem has been to determine the values of the constants which appear in the equation. But a very different class of cases now demands a passing notice.

A series of observations is sometimes found to be affected with errors too great to be explained as the result of unavoidable and fortuitous causes, and it becomes apparent that the law of recurrence of these errors must be determined before the observations can be made to yield any valuable results. The American parties which were sent out in 1874 to observe the transit of Venus were provided with instruments for the determination of their local time, of such a character that the accidental error of a determination from a single star might fairly be estimated at $0^s.05$ or $0^s.06$, but results obtained from observations of different stars varied among themselves by more than ten times this amount. An inspection of the discrepancies having shown that they depended in some way upon the distance of the observed star from the zenith, it was found by trial that the error at any zenith distance, z, could be represented by the expression

$$E = \pm \{a \cos z - b \sin 2z\}$$

where a and b are constants whose values were found from the observations themselves. The physical cause of these errors was subsequently found to be the bending of the instrument under its own weight, but it is to be noted that the above law of recurrence of the errors was determined first, the cause afterwards.

Expressions of this kind are sometimes called *interpolation formulæ* and sometimes *empirical equations;* the one term having reference to their use, the other to their derivation. They are of very general use in all branches of physical science, since they may be made to serve as a convenient summary of a vast amount of numerical data, and one of the most important applications of the method of least squares is in determining the values of the constants which enter into such expressions. The problems treated in §§ 1 and 10 both belong to this class, and the following expression for the magnetic declination at Washington, D.C., derived by *Mr. C. A. Schott* * from a series of observations extending over ninety years may serve as a further illustration:

$$\text{Mag. Dec.} = 2°.47 + 2°.50 \sin [1°.40 (\mathsf{T} - 1850) - 14°.6]$$

where T denotes the year for which the declination is required.

When the cause whose effects are to be represented by an equation is known, the form of this equation can usually be derived by mathematical analysis; but where empirical formulæ are employed other methods must be resorted to. The simplest of these is a graphical representation of the errors or other data under consideration. For this purpose let the errors represent ordinates, and the values of any variable upon which they are supposed to depend, the corresponding abscissas. Let points be plotted with these ordinates and abscissas as was done in obtaining the form of the error curve, Figs. A, B, C, D, and let a smooth curve be drawn through these points either free-hand or by the aid of a draughtsman's "irregular curve." The distance of each plotted point from the curve, *measured along an ordinate,* is the residual corresponding to the point,

* *U. S. C. & G. S. Report,* 1882, p. 258.

and in accordance with the principle of least squares the curve should be so drawn as to make the sum of the squares of these residuals as small as possible, without unduly complicating the curve. If the variable has been properly chosen it will in many cases be found possible to draw a simple curve which shall represent the data within the limits of the accidental errors of observation, and as this curve is the graphical representation of the law required, its equation, $y = f(x)$, is the analytical representation of that law. In this manner the form of the equation treated in § 10 was obtained.

In other cases it will be possible to draw a smooth and simple curve which shall not represent the data within the limits of accidental error of the observations, but about which the points will be grouped, alternating from one side to the other in a systematic manner. Let the excess of the ordinate of any point over the corresponding ordinate of the curve be plotted with the given abscissa in a new curve. The two curves thus constructed will together form the graphical representation of the law of the data, and the analytic expression of that law will be

$$y = f(x) + \phi(x)$$

if $y = f(x)$ and $y = \phi(x)$ are the equations of the two curves respectively.

In some cases the curves themselves will be a sufficient representation of the data, and it will be unnecessary to determine their equations since the value of y corresponding to any given value of x may be obtained by direct measurement. In other cases the curve will be chiefly serviceable in suggesting the probable form of an equation between the observed quantity and a variable upon which it is supposed to depend, or in showing that no simple relation exists between them. Two forms of equation are of such frequent use in this connection that they deserve especial notice.

If the plotted curve does not differ very greatly from a straight line, the relation of the variable x to its function y may be represented by

$$y = a + bx + cx^2 + dx^3 + \text{etc.} \qquad (35)$$

This equation contains the first few terms of an infinite series by which a limited arc of any continuous curve can be represented, and since the actual relation between y and x could be represented by a curve, if its mathematical expression were known, it follows that the above equation can be made to represent this relation over a certain range of values of x, by assigning proper values to the coefficients. The number of terms of this series which should be taken into account, and the limits of x beyond which the equation is not applicable, depend upon the actual relation between y and x, and are therefore unknown; but, in general, it is not well to attempt to use this equation for large values of x, or when more than three or at most four terms are required. Its application in a simple case is illustrated in the problem of § 1, where y and x being replaced by the length of the bar and its temperature, it is assumed that their relation can be expressed within the range of temperature over which the observations extend, by the first two terms of the series.

The second type of equation above referred to is

$$\left.\begin{aligned} y = a_0 + a_1 \cos\frac{x}{m} + a_2 \cos\frac{2x}{m} + a_3 \cos\frac{3x}{m} + \text{etc.} \\ + b_1 \sin\frac{x}{m} + b_2 \sin\frac{2x}{m} + b_3 \sin\frac{3x}{m} + \text{etc.} \end{aligned}\right\} \quad (36)$$

in which m is an undetermined constant expressed in the same unit as x; $\frac{x}{m}$ is therefore a ratio, or absolute number, whicl in the application of the equation to numerical data must be transformed into circular measure by multiplying it by $\frac{180^\circ}{\pi} = 57^\circ.29578$. This form of equation may be made to represent any relation whatever between finite values of y and x, including those cases in which y is a discontinuous function, but it is especially advantageous when y is a periodic function of x, i.e., one in which the same values of y recur for values of x, separated by a constant interval, τ, called the period of x, so that

$$f(x) = f(x+\tau) = f(x+2\tau) = \cdots = f(x+n\tau).$$

The simplest type of such a function is $y = \sin x$, the period in this case being $\tau = 360°$; e.g.,

$$\sin 10° = \sin (10° + 360°) = \sin (10° + 720°) = \text{etc.}$$

When y is such a function the constant m should be put equal to the period divided by 2π, $m = \frac{\tau}{2\pi}$; in other cases the value of m should be so taken that the greatest value of $\frac{x}{m}$ included among the data shall not exceed π. The application of this formula may often be facilitated by noting that if the relation between y and x is such that $f(+x) = f(-x)$, the sine terms all disappear, and the equation reduces to

$$y = a_0 + a_1 \cos\frac{x}{m} + a_2 \cos\frac{2x}{m} + \text{etc.} \qquad (37)$$

while if $f(x) = -f(-x)$, the cosine terms vanish, and the equation becomes

$$y = b_0 + b_1 \sin\frac{x}{m} + b_2 \sin\frac{2x}{m} + \text{etc.} \qquad (38)$$

The several forms above given to this type of equation are those most convenient for use when the values of the coefficients a, b, etc., are to be determined, but after their numerical values have been found it is advantageous to transform the equation as follows: Introduce the auxiliary quantities

$$n_0,\ n_1,\ n_2,\ N_1,\ N_2,\ \text{etc.}$$

defined by the relations

$$n_0 = a_0 \qquad n_1 \cos N_1 = a_1 \qquad n_2 \cos N_2 = a_2$$
$$n_1 \sin N_1 = b_1 \qquad n_2 \sin N_2 = b_2$$

and the equation becomes

$$y = n_0 + n_1 \cos\left(\frac{x}{m} - N_1\right) + n_2 \cos\left(\frac{2x}{m} - N_2\right) + \text{etc.} \qquad (39)$$

each pair of terms of the original equation being here replaced by a single term. The expression for the magnetic declination

at Washington given on p. 59 is of this type, as may be seen by writing it in the equivalent form

$$\text{Mag. Dec.} = 2°.47 + 2°.50 \cos [1°.40(\text{T}-1850) - 104°.6]$$

The mode of applying this form of equation may be illustrated by means of the following data selected from the series of observations whose residuals are plotted in Fig. C. The observed quantity, B, is the difference of stellar magnitude (brightness) between the planet Saturn and his satellite Iapetus. The quantity l given with each observed B fixes the position of the satellite in its orbit at the time of observation, and is analogous to the variable angle θ in a system of polar coordinates.

l.	B.	Residual.	l.	B.	Residual.
	m	m		m	
10°	10.82	− 0.28	200°	10.66	+ 0.28
70	11.81	+ 0.22	230	9.87	− 0.21
110	11.69	− 0.02	270	10.43	+ 0.21
140	11.42	− 0.03	310	10.48	− 0.15

B is here seen to run through a complete cycle of values between the limits 9.87 and 11.81, while l varies from 0° to 360°. We shall therefore endeavor to represent B as a periodic function of l whose period is 360°. In accordance with this assumption we put

$$\frac{x}{m} = \frac{l}{\frac{360°}{2\pi}} \cdot \frac{180°}{\pi} = l; \; y = B$$

and taking into account the first five terms of the series, the several observations furnish the following

Observation Equations.

$$10.82 = a_0 + 0.98\,a_1 + 0.17\,b_1 + 0.94\,a_2 + 0.34\,b_2$$
$$11.81 = a_0 + 0.34\,a_1 + 0.94\,b_1 - 0.77\,a_2 + 0.64\,b_2$$
$$11.69 = a_0 - 0.34\,a_1 + 0.94\,b_1 - 0.77\,a_2 - 0.64\,b_2$$
$$11.42 = a_0 - 0.77\,a_1 + 0.64\,b_1 + 0.17\,a_2 - 0.98\,b_2$$
$$10.66 = a_0 - 0.94\,a_1 - 0.34\,b_1 + 0.77\,a_2 + 0.64\,b_2$$

$$9.87 = a_0 - 0.64\,a_1 - 0.77\,b_1 - 0.17\,a_2 + 0.98\,b_2$$
$$10.43 = a_0 + 0.00\,a_1 - 1.00\,b_1 - 1.00\,a_2 + 0.00\,b_2$$
$$10.48 = a_0 + 0.64\,a_1 - 0.77\,b_1 - 0.17\,a_2 - 0.98\,b_2$$

The solution of these equations will be found in the following section. The values obtained for the constants are

$$a_0 = 10.92,\quad a_1 = +0.15,\quad b_1 = +0.74,\quad a_2 = -0.04,\quad b_2 = -0.18$$

Introducing the constants n, N, and determining their values from the relations

$$n_0 = 10.92 \qquad n_1 \cos N_1 = +0.15 \qquad n_2 \cos N_2 = -0.04$$
$$n_1 \sin N_1 = +0.74 \qquad n_2 \sin N_2 = -0.18$$

the equation becomes

$$B = 10.92 + 0.76 \cos(l - 78°.5) - 0.18 \cos(2l - 77°.5)$$

The residuals obtained by comparing the values of B computed from this formula with the observed values are given above with the data.

Abundant data for exercise in deriving empirical formulæ of this kind may be found in the *United States Coast and Geodetic Survey Report* for 1882, pp. 218–257.

§ **16. Approximate Solutions.** It is often desirable to obtain from a series of observations, as rapidly and with as little labor as possible, a set of values of the unknown quantities involved which shall be fair approximations to their most probable values, but in which the highest degree of accuracy is not required. In cases of this kind the least square treatment of the observation equations as illustrated in § 10 is too long and laborious, and the following method may often be substituted for it with advantage.

Let there be any number, e.g., three, unknown quantities involved in a set of observation equations of the form

$$ax + by + cz + n = 0 \qquad \text{Weight} = p$$

and let each of these equations be multiplied by its weight, giving the group of equations,

$$\left.\begin{array}{llll} a_1x+b_1y+c_1z+n_1=0 & & & k_1 \\ a_2x+b_2y+c_2z+n_2=0 & & & k_2 \\ \text{etc.} & \text{etc.} & \text{etc.} & \text{etc.} \end{array}\right\} \tag{40}$$

Multiply each of these equations by the undetermined constant k placed opposite it, and let the sum of all the resulting equations be formed. By the use of the summation symbol, [ ], this sum may be written

$$[ka]x+[kb]y+[kc]z+[kn]=0$$

Since the several values of k which enter into this equation are entirely arbitrary it would be permissible to assign to them such values that $[kb]$ and $[kc]$ should each equal 0, which would give at once

$$x=-\frac{[kn]}{[ka]} \tag{41}$$

This, however, is not practically advantageous on account of the labor involved in determining the values of k. We therefore put

$$x=-\frac{[kn]}{[ka]}-\frac{[kb]}{[ka]}y-\frac{[kc]}{[ka]}z \tag{42}$$

and, limiting the values of k to $+1$ and -1, assign them in such a manner that $[ka]$ shall be made as great, and $[kb]$, $[kc]$ as small as possible. In this manner the coefficients of y and z may often be made so small that if approximate values of y and z are substituted in equation (42), they will furnish a sufficiently accurate value of x; since the effect of the errors of these approximations will be much diminished by the small coefficients by which they are multiplied.

The value of y may be found in the same manner by selecting a set of k's which shall make $[kb]$ large and $[ka]$, $[kc]$ small, and similarly for z. Two or three trials may be required before sufficiently close approximations to the values of x, y, z are obtained, but these trials are rapidly and easily made, and, if necessary, in exceptional cases the summation equations may be written in the form

$$\left.\begin{aligned} [k'a\ ]x+[k'b\ ]y+[k'c\ ]z+[k'n\ ]=0 \\ [k''a\ ]x+[k''b\ ]y+[k''c\ ]z+[k''n\ ]=0 \\ [k'''a]x+[k'''b]y+[k'''c]z+[k'''n]=0 \end{aligned}\right\} \qquad (43)$$

and the equations solved by any of the methods of elementary algebra, but in every case the values of $k+1$ and -1 should be so chosen that in the first equation the coefficient of x, in the second equation the coefficient of y, and in the third equation the coefficient of z, shall be made as large, and all of the other coefficients as small as possible.

By this mode of solution each observation with its proper weight is included in the determination of the unknowns, but since the principle of least squares has not been taken into account, it cannot be expected that the resulting values will be the best that the observations can be made to yield.

To illustrate the mode of solution we recur to the observation equations contained in the preceding section and putting $a_0 = 10.00 + a$ write them as follows, placing the several values of k at the right of each equation.

$\cos l$	$\sin l$	$\cos 2l$	$\sin 2l$	k'	k''	k'''	k^{iv}	k^{v}
$0.82 = a + 0.98\,a_1 + 0.17\,b_1 + 0.94\,a_2 + 0.34\,b_2$				+1	+1	+1	+1	+1
$1.81 = a + 0.34\,a_1 + 0.94\,b_1 - 0.77\,a_2 + 0.64\,b_2$				+1	+1	−1	−1	+1
$1.69 = a - 0.34\,a_1 + 0.94\,b_1 - 0.77\,a_2 - 0.64\,b_2$				+1	−1	+1	−1	−1
$1.42 = a - 0.77\,a_1 + 0.64\,b_1 + 0.17\,a_2 - 0.98\,b_2$				+1	−1	+1	+1	−1
$0.66 = a - 0.94\,a_1 - 0.34\,b_1 + 0.77\,a_2 + 0.64\,b_2$				+1	−1	−1	+1	+1
$-0.13 = a - 0.64\,a_1 - 0.77\,b_1 - 0.17\,a_2 + 0.98\,b_2$				+1	−1	−1	−1	+1
$0.43 = a + 0.00\,a_1 - 1.00\,b_1 - 1.00\,a_2 + 0.00\,b_2$				+1	+1	−1	−1	−1
$0.48 = a + 0.64\,a_1 - 0.77\,b_1 - 0.17\,a_2 - 0.98\,b_2$				+1	+1	−1	+1	−1

The summation equations obtained from this group are:

$$\left.\begin{aligned} +7.18 &= 8\,a - 0.73\,a_1 - 0.19\,b_1 - 1.00\,a_2 + 0.00\,b_2 \\ -0.10 &= 0\,a + 4.65\,a_1 - 1.13\,b_1 - 1.00\,a_2 + 0.00\,b_2 \\ +4.30 &= 0\,a + 1.15\,a_1 + 5.57\,b_1 + 0.14\,a_2 + 0.00\,b_2 \\ -0.42 &= 0\,a + 0.53\,a_1 - 0.41\,b_1 + 4.42\,a_2 + 0.00\,b_2 \\ -0.86 &= 0\,a + 0.21\,a_1 + 0.19\,b_1 + 2.54\,a_2 + 5.20\,b_2 \end{aligned}\right\} \qquad (44)$$

and these correspond to the normal equations of a least square solution. To apply the method of approximations to the solution of this group of equations, we write them in the form:

$$\left.\begin{aligned} \alpha &= +0.897 + 0.091\,a_1 + 0.024\,b_1 + 0.125\,a_2 \\ a_1 &= -0.022 + 0.000\,a_1 + 0.243\,b_1 + 0.215\,a_2 \\ b_1 &= +0.772 - 0.206\,a_1 + 0.000\,b_1 - 0.025\,a_2 \\ a_2 &= -0.095 - 0.120\,a_1 + 0.093\,b_1 + 0.000\,a_2 \\ b_2 &= -0.166 - 0.040\,a_1 - 0.036\,b_1 - 0.488\,a_2 \end{aligned}\right\} \qquad (45)$$

The divisions required in making this transformation were made by the use of *Crelle's Rechentafeln.*

By operations which can be performed mentally, we obtain the following sets of approximations to the values of

	I.	II.	III.
a_1	0.0	+0.2	+0.15
b_1	+0.8	+0.7	+0.74
b_2	−0.1	−0.0	−0.04

and substituting in equations (45) the values given under III. we find the adopted values

$$a_0 = 10 + \alpha = 10.92 \qquad a_1 = +0.15 \qquad a_2 = -0.04$$
$$b_1 = +0.74 \qquad b_2 = -0.18$$

which were employed in the preceding section.

INDEX TO FORMULÆ.

$y = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$ . . . . . § 3, 4

$\frac{1}{2h^2} = \frac{[vv]}{n}$ . . . . . § 5

$x_0 = \frac{[pm]}{[p]}$ . . . . . § 6

$[aa]x + [ab]y + [ac]z + [an] = 0$ . . . . . § 7

$[ab] = \frac{1}{2}\{[(a+b)^2] - ([aa]+[bb])\}$ . . . . . § 9

$[as] = [aa]+[ab]+[ac]+\cdots+[an]$ . . . . . § 9

$[bc\cdot 1] = [bc] - \frac{[ab]}{[aa]}[ac]$ . . . . . § 9

$[cn\cdot 2] = [cn\cdot 1] - \frac{[bc\cdot 1]}{[bb\cdot 1]}[bn\cdot 1]$ . . . . . § 9

$[aa]k_1 + [a\beta]k_2 + \phi(x_0, y_0, z_0) = 0$ . . . . . § 11

$r = \frac{0.477}{h} = \pm 0.674\sqrt{\frac{[vv]}{n-\mu}} = \pm 0.845\frac{[+v]}{\sqrt{n(n-\mu)}}$ . . § 12

$r_0 = \frac{r}{\sqrt{n}} = \frac{r_1}{\sqrt{[p]}}$ . . . . . § 13

$p' : p'' : p''' = \frac{1}{r'^2} : \frac{1}{r''^2} : \frac{1}{r'''^2}$ . . . . . § 14

$y = a + bx + cx^2 + \text{etc.}$ . . . . . § 15

$y = n_0 + n_1\cos\left(\frac{x}{m} - N_1\right) + n_2\cos\left(\frac{2x}{m} - N_2\right) + \text{etc.}$ . . § 15

$-x = \frac{[kn]}{[ka]} + \frac{[kb]}{[ka]}y + \frac{[kc]}{[ka]}z \qquad k = \pm 1$ . . . . § 16

REESE LIBRARY OF THE UNIVERSITY OF CALIFORNIA

UNIVERSITY OF CALIFORNIA LIBRARY
BERKELEY

Return to desk from which borrowed.
This book is DUE on the last date stamped below.

15Feb'51WK
11Aug52FA
JUL 28 1952 LU
14Dec'52 CB
JAN 29 1953
19Mar'53LO
MAR 7 1953 LU
25Apr'58KT
JUN 10 1953 LU

LIBRARY USE
AUG 27 1959
REC'D LD
AUG 27 1959
4Nov'64SB
REC'D LD
OCT 28 '64 -6 PM

LD 21-100m-11,'49(B7146s16)476